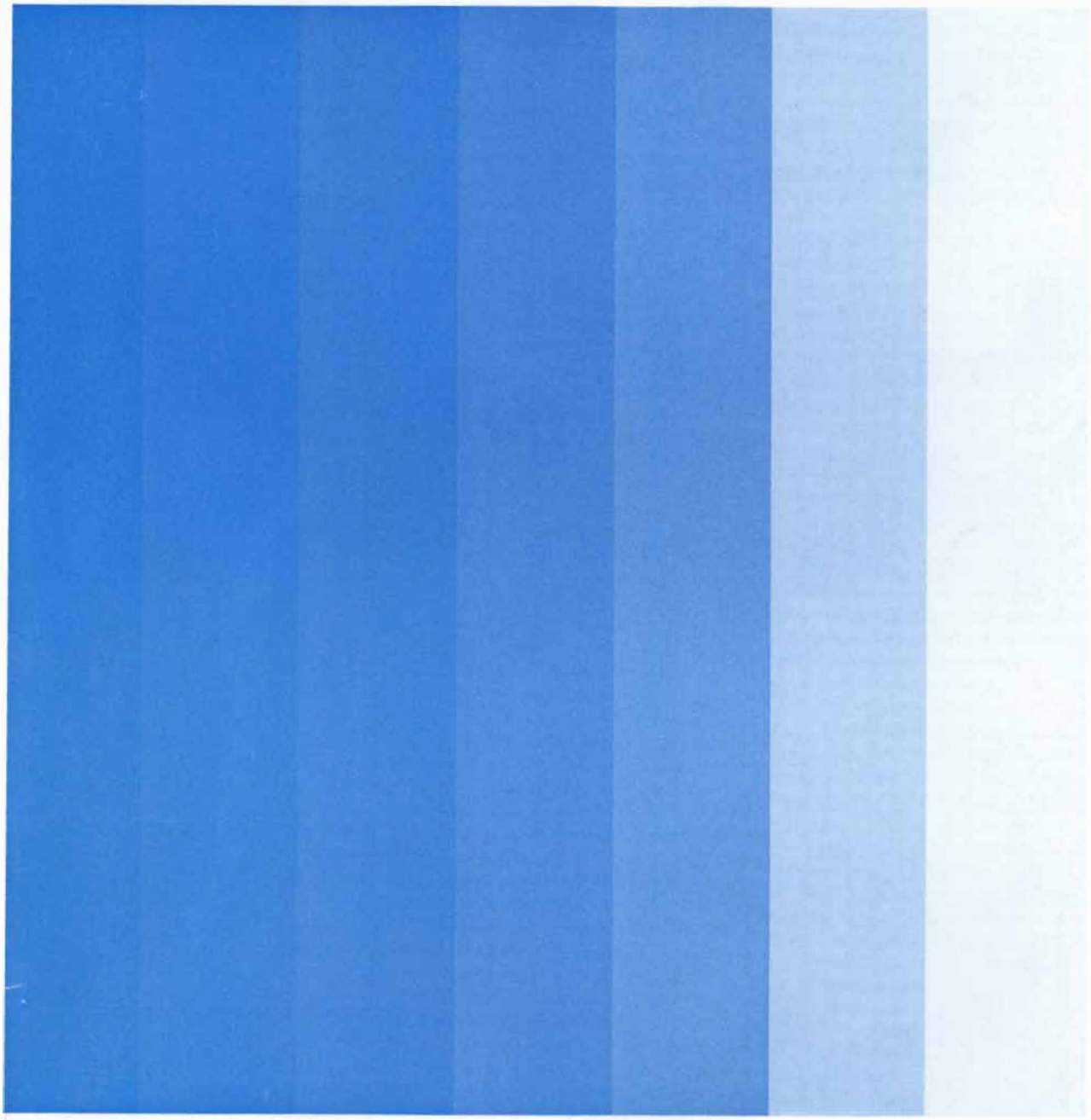
SCIENTIFIC AMERICAN



CONTOUR AND CONTRAST

ONE DOLLAR

June 1972

Variations on a basic Pinto.

Ford Pinto Wagon: Rugged 2,000cc engine, front disc brakes and over 60 cubic feet of cargo space. (VW Squareback and Vega Kammback have only about 50.) The picture shows the Squire Option, plus luggage rack, whitewall tires and bumper guards.

Fresh orange, anyone? Pinto's Sports Accent group has bright new color accents like orange or avocado, with matching interiors and vinyl roofs. And steel-belted radial-ply white sidewall tires for long mileage.

The Sunshine Pinto: Our new easy-opening, solid-closing sun roof comes at a remarkably low price. It's shown above on a stylish little 3-Door Runabout, with vinyl roof and luxury decor options.

Red, white and blue! The new Sprint Option includes this racy color scheme, inside and out. Plus dual racing mirrors and white sidewall tires.

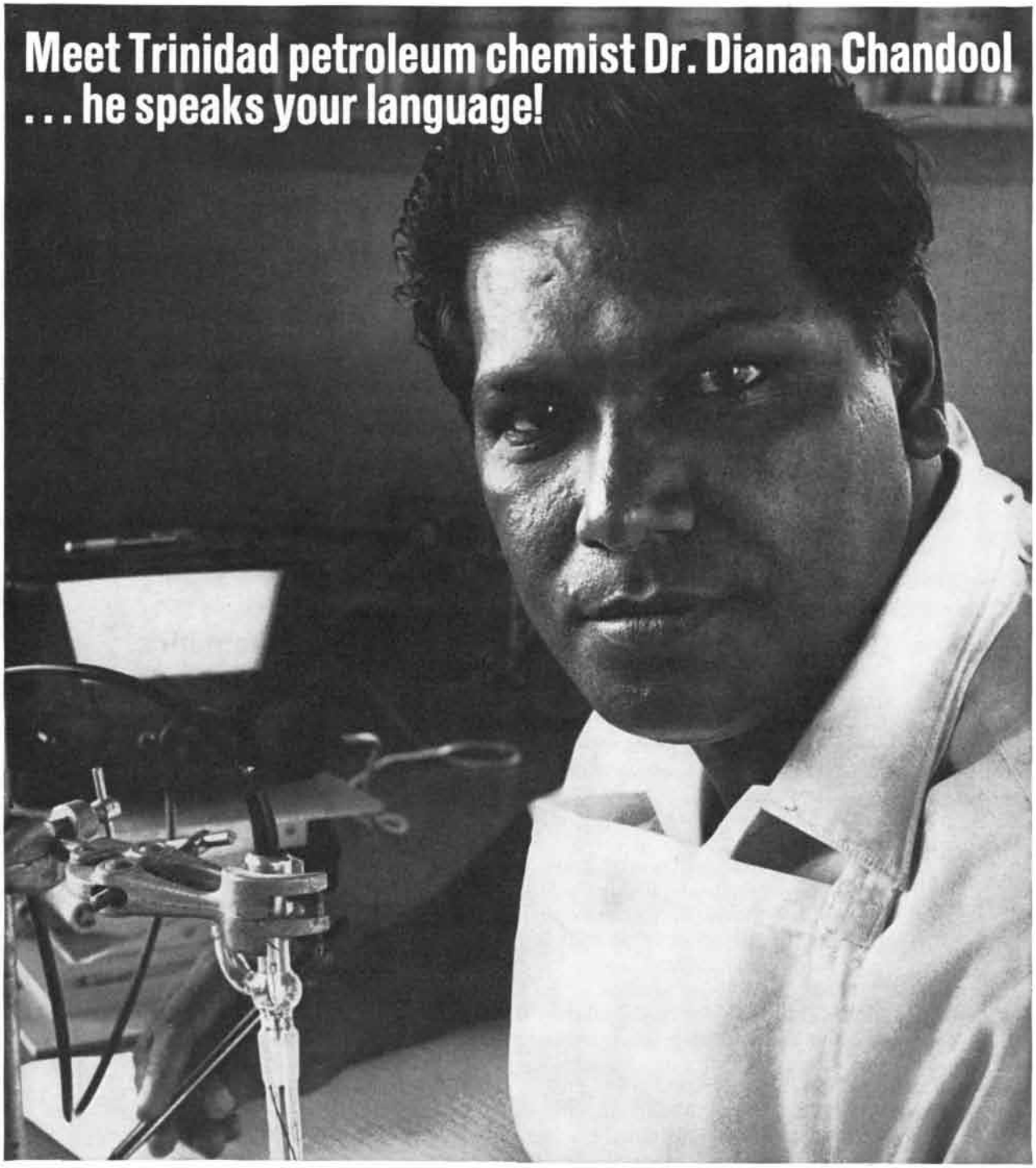
Ford Pinto comes in a lot of different variations. But the basic theme is always the same: solid, sensible, and economical.

When you get back to basics, you get back to Ford.

FORD PINTO

FORD DIVISION Ford

Meet Trinidad petroleum chemist Dr. Dianan Chandool . . . he speaks your language!



Finding people in a developing country, that's easy. But finding trained people with combined administrative and technological expertise, that's a different story.

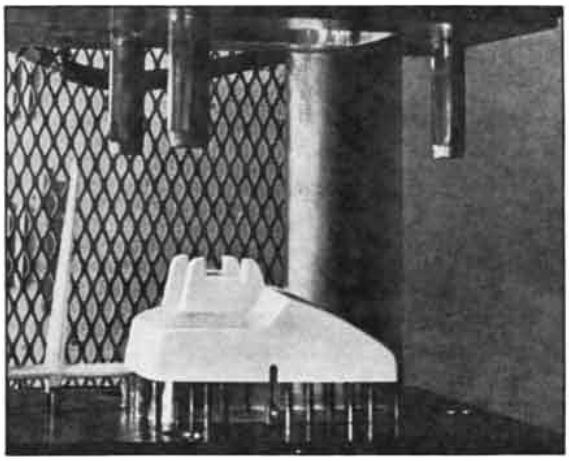
You're looking at one of them, Dr. Dianan Earle Chandool. He heads agricultural chemical development for Shell Trinidad, Ltd. Dr. Chandool likes what he's doing and how it fits into his country's plan for progress. "I believe through my job," he says, "I am helping in a small way to build our nation."

Pride, professionalism, patriotism . . . those are good qualities for any employee or partner. You find them over and over again in Trinidad & Tobago. Then add to the right people the right kind of social and political climate, strong industrial incentives, central market position: you've got a pretty good combination. We can tell you more about it, after you fill in this coupon.

Trinidad & Tobago
Industrial Development Corp.
400 Madison Avenue, New York, N. Y. 10017
(212) PLaza 5-9360
Clyde S. Namsoo: North American Director

Please send me factual information on Trinidad & Tobago as an investment location.

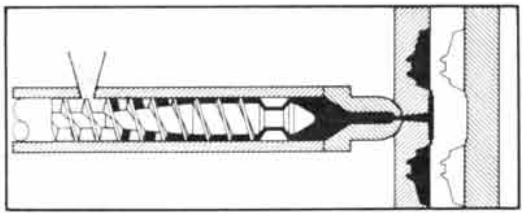
Name & Title _____
Company _____
Address _____ Zip _____
City & State _____ Tel. No. _____
Product Line _____



Molding by the millions. Western Electric people produce some 8 million phones a year. Molded plastic is used for housings and many other parts. So there is a constant investigation into the most effective way to use these materials.

$$A^*(z,t) = A_e^*(z) - [(A_f^* - A_i^*)/(1 - e^{-\beta N t_r})]e^{-\beta N t}$$

In developing the model at Western Electric's Engineering Research Center, it was found that melting behavior can be described by this formula which includes terms for shear heating and conduction heating effects. Other models were developed for temperature and pressure profiles.



End of molding cycle. At this point, the screw is stationary and heat is conducted into the plastic on the screw. After the plastic solidifies, the mold is opened as shown. The parts can then be ejected.

Solving the mysteries of molding with mathematics.

Even though plastics have been around for many years, there's still a lot to be learned about these versatile materials and their processing. So they are the subject for continuing studies by our engineers.

Some of their recent investigations have brought forth new and highly useful information about a relatively unexplored area: the melting behavior of plastics in the injection molding process.

One result of these studies is the mathematical formula, or model, above.

The model helps us predict melting behavior along the length of the injection screw molding machine used to mold telephone housings and other parts. Melting behavior is extremely important, because plastic pellets should be completely melted but not thermally decomposed before injection into the mold.

This information on melting is then used to investigate screw designs, operating conditions, machine sizes and plastic properties. All of which is aimed at obtaining optimum processing techniques.

Predictions obtained from the mathematical model have checked out closely with experimental observations. So the resulting screw designs are now undergoing evaluation by engineers at our plants in Indianapolis and Shreveport.

Conclusion: For new designs and materials, the model can help reduce the development cost for new molded parts and materials. For manufacturing current products, operating costs can be reduced.

Perhaps most significant is that we're getting information about molding temperatures not available experimentally. And many other types of information can be obtained without the use of costly, time-consuming experimental work.

The end result will be more efficient plastic molding and therefore a better product for the lowest possible cost.

Western Electric

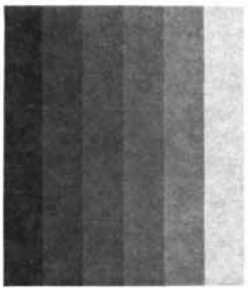
We make things that bring people closer.

Established 1845 **SCIENTIFIC AMERICAN** June 1972 Volume 226 Number 6

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $10 PER YEAR. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 128, BORGERHOUT 2200, BELGIUM.



THE COVER

The illustration on the cover is an example of a visual illusion known as border contrast. Each vertical band in the illustration appears to be lighter in color on its left side than on its right side. Each band, however, is of an even shade from one edge to the other. The reader can demonstrate this for himself by placing two sheets of paper over the illustration so that only one band is visible. Moreover, by slowly drawing the two covering sheets of paper apart, the reader can determine how many steps in contrast are required to produce the strongest border effect. An explanation of the neural mechanisms that are responsible for this effect and several others is given by Floyd Ratliff in his article "Contour and Contrast" (*page 90*).

THE ILLUSTRATIONS

Page	Source
15	Ilil Arbel
16–21	Allen Beechel
22–26	U.S. Navy
27	Ilil Arbel
29–33	Bunji Tagawa
34	Barry D. Kahan, Massachusetts General Hospital (*top*); Bunji Tagawa (*bottom*)
35	Barry D. Kahan, Massachusetts General Hospital
36–37	Bunji Tagawa
38	Jon Brenneis
39–46	Andrew Tomko
55	P. E. Playford (*top*), John S. Shelton (*bottom*)
56–61	Tom Prentiss
63	Ilil Arbel
64	Tom Prentiss
70	Treat Davidson, from National Audubon Society
72–74	Lorelle M. Raboni
75–77	Bernd Heinrich, University of California at Berkeley, and George A. Bartholomew, University of California at Los Angeles
79–84	Ilil Arbel
90	Kröller-Müller Museum
92	Jim Egleson (*top and bottom*); Floyd Ratliff, Rockefeller University (*middle*)
93	Jim Egleson (*left*); Floyd Ratliff, Rockefeller University (*right*)
94	Jim Egleson
95	Jay M. Enoch, Washington University School of Medicine (*top*); Floyd Ratliff, Rockefeller University (*bottom*)
96–97	Ben Rose
98	Jim Egleson
99	Private collection, New York
100–101	Floyd Ratliff
104–106	Eric Mose
107–108	Ilil Arbel
109–110	Dan Todd
112–117	Ilil Arbel
123–128	Roger Hayward

Buckle up for safety. Pontiac Motor Division

Choose your tow car as carefully as you choose your trailer.

Ever notice how it works? As the scenery gets nicer, the roads get tougher. Which is why you want your tow car as strong and roadworthy as your trailer.

After all, if your car doesn't have enough equipment to handle your trailer, you're risking trouble. If it has too much equipment, you're wasting money.

How do you get a car that's just right for your trailer? Well, you can sit down and try to figure out all the things you need, then have them installed on your present car. But that gets confusing, time-consuming and expensive.

It makes more sense to order a car with the correct equipment right from the factory. And that's where Pontiac can really help.

You see, Pontiac has over ten years' experience in designing trailer-towing packages for cars.

That experience has led to the development of seven different towing packages. Each with the right combination of springs, shocks, tires, wheels, cooling gear and power to do a specific kind of towing job.

So with Pontiac, there's no guessing or waiting around. You pick the towing package that matches the Pontiac you'd like to drive and the load you want to haul. It's that simple.

Which package goes on which Pontiac to haul which trailer? Pontiac's 1972 Trailer-towing Catalog has all the information. Get a free copy by writing: Trailer Towing, Dept. 72S2, Pontiac Motor Division, One Pontiac Plaza, Pontiac, Michigan 48055.

Wide-Track Pontiac ...a cut above.

Name _____

Address _____

City _____

State _____ Zip _____

Remember to Also Ask for a Free Copy of Pontiac's 1972 Directory of National Camp Grounds.

LETTERS

Sirs:

Donald P. Snowden is to be congratulated on his timely article "Superconductors for Power Transmission" [SCIENTIFIC AMERICAN, April]. There is no doubt that formidable problems await the introduction of a technology as complicated as cryogenics into another as complicated as power generation and transmission. However, the increasing demand for power coupled with rising concern over the unsightliness of overhead lines and towers will force this union within the next two decades. A group at the Brookhaven National Laboratory has recently completed a study for the National Science Foundation on the application of superconductivity to underground power transmission. We concur with many of the statements made, particularly the use of niobium-tin in future superconducting cables, although the temperature-independent loss up to 12 Kelvin claimed by Snowden is controversial. This may be a unique property of the concentric-layer construction described and, if so, represents an important development. We based our cable calculations on loss measurements of commercially available single-layer niobium-tin ribbon and achieved satisfactory designs: a hyperbolic increase of losses with temperature was assumed. We would also dispute the conclusion that either helium or vacuum will form the insulating dielectric; the breakdown behavior of bulk materials under very-high-voltage stress is inferior to a laminar form of construction, and our studies led us to believe that an impregnated wrapped tape would present less chance of voltage breakdown and would still possess acceptable dielectric losses.

The point made that this class of cables would be rated to match the average load and not the peak load, as is now the case with conventional cables, is very important. The regional loads in the U.S. are becoming characterized by marked diurnal variations. Another beneficial aspect of superconducting cables that may be mentioned is that the electrical characteristics closely resemble overhead lines and thus underground networks many hundreds of miles in extent are technically, if not economically, feasible. This is not the case with conventional underground cables, which must be limited to a few tens of miles. Finally, let us point out that superconducting cables will contain a negligible volume of conductor and on a weight-for-weight basis can be expected to carry about 20 times as much power as a conventional cable. This suggests that long lengths of flexible cable should not be too difficult to install.

E. B. FORSYTH

M. GARBER

J. E. JENSEN

G. H. MORGAN

Brookhaven National Laboratory
Upton, N.Y.

Sirs:

The comments of the Brookhaven group give further emphasis and support for a major point of my article: the need and the viability of superconducting power transmission in the next decades.

Scientific American, June, 1972; Vol. 226, No. 6. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Manuscripts are submitted at the author's risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 25 cents.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 3578.

Please enter change of address ☐
Please enter subscription ☐
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25

NAME

NEW ADDRESS

OLD ADDRESS

I should like to comment specifically on two points raised in their letter. The loss results on niobium-tin described in my article are controversial only because the values are agreeably low (nature is often not this generous) and because they have been determined thus far in only one laboratory on small specimens. I know of no measurements on samples of a similar geometry that indicate the data presented to be incorrect. Certainly in the orderly development of the field of superconducting power transmission it will be necessary to show that these loss values can be obtained in specimens approximating the sizes that are required in actual transmission systems.

I agree that the question of the choice of dielectric is far from settled. Considerably more research must be carried out to determine the best choice. Indeed, at this time the amount of work required for the dielectric system probably exceeds that for the superconductor itself. It is interesting to note that the same question is also far from settled for cryogenic, nonsuperconducting systems. Both vacuum and wrapped, impregnated-tape insulations for cryogenic lines have had their strong supporters for many years and both systems are currently being actively developed. The choice of rigid or flexible conductors is also strongly influenced by the choice of dielectric. Experience obtained with cryogenic systems in the next few years may help to clarify this question for superconducting lines.

DONALD P. SNOWDEN

Gulf General Atomic Company
San Diego, Calif.

ERRATA

In the article "The Cratering of Indochina," by Arthur H. Westing and E. W. Pfeiffer (SCIENTIFIC AMERICAN, May), the Roman numerals designating the four military regions of South Vietnam in the map on page 22 were incorrectly given in reverse order; they should read in ascending order from north to south, not south to north as shown. In the bar chart at the bottom of page 24 the legends "POUNDS PER PERSON" and "POUNDS PER ACRE" should be transposed.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1922: "A reconvention of the National Radio Conference is being held at this writing and will continue until all the matters pertaining to radio regulation are reduced to a workable and satisfactory basis. The general opinion is that radio communication is a public utility and as such should be regulated and controlled by the Federal Government in the public interest. One recommendation is that the same wavelength (or overlapping bands) not be assigned to stations within a certain distance from one another, and that the National Bureau of Standards should make a study of the width of the wave band required for satisfactory radio telephony."

"At the meeting of the British Association in Edinburgh in September last a joint discussion on the age of the earth took place under the auspices of the sections for mathematics, geology, zoology and botany. The inadequacy of Lord Kelvin's original estimate of 20 or 30 million years has long been admitted. The more recent estimates are for a much greater age. Lord Rayleigh considered that the most accurate estimate of the age of the earth can be derived from the rate of radioactive disintegration. Professor Gregory dealt with the geological estimate of the age of the earth, based on the salinity of the sea. On the whole, by these and other methods there is now a satisfactory agreement between the results of arguments based on astronomical, physical and geological considerations. These indicate an age of the earth, since solidification, of 1,000 million years."

"Recognition is being given these days to the invisible portion of the spectrum, which is known to exceed greatly in range the portion to which the human eye is sensitive. Certain animalcules that will survive exposure to violet light for four or five hours are killed by the ultraviolet within 15 seconds. The effect of these rays on the human skin, eye and so on is different only in degree and in no sense in kind. It therefore becomes an object to screen off these rays from ordinary light, if it can be done. This hope is realized in a substance that is now being marketed in England after extensive tests by numerous prominent scientists. It is anticipated that this substance, whatever it is, will play an important part in the future development of the optical industry."

"Many SCIENTIFIC AMERICAN stories can well be retold in motion picture form, for the reason that they require animation for a readier understanding by the layman. To this end we have entered the motion picture field, as producers of SCIENTIFIC AMERICAN films, in collaboration with the Coronet Films Corporation of Providence, R.I. Such subjects as can be treated to the best advantage in motion picture form will be taken from our columns and transplanted to the screen."

SCIENTIFIC AMERICAN

JUNE, 1872: "The progress of the movement to the eight-hour working day from the 10-hour day, which until recently appeared successful, has encountered a check that bids fair to result in defeat. Elated by the easy victories gained over the smaller employers, the strikers have carried the war to the doors of the great manufacturing firms and corporations. But here a strong opposition has been encountered. The eight-hour movement affects every working man in the land, and unless all or a very large majority of the laboring classes give it unwavering support, the accomplishment of its design is impracticable."

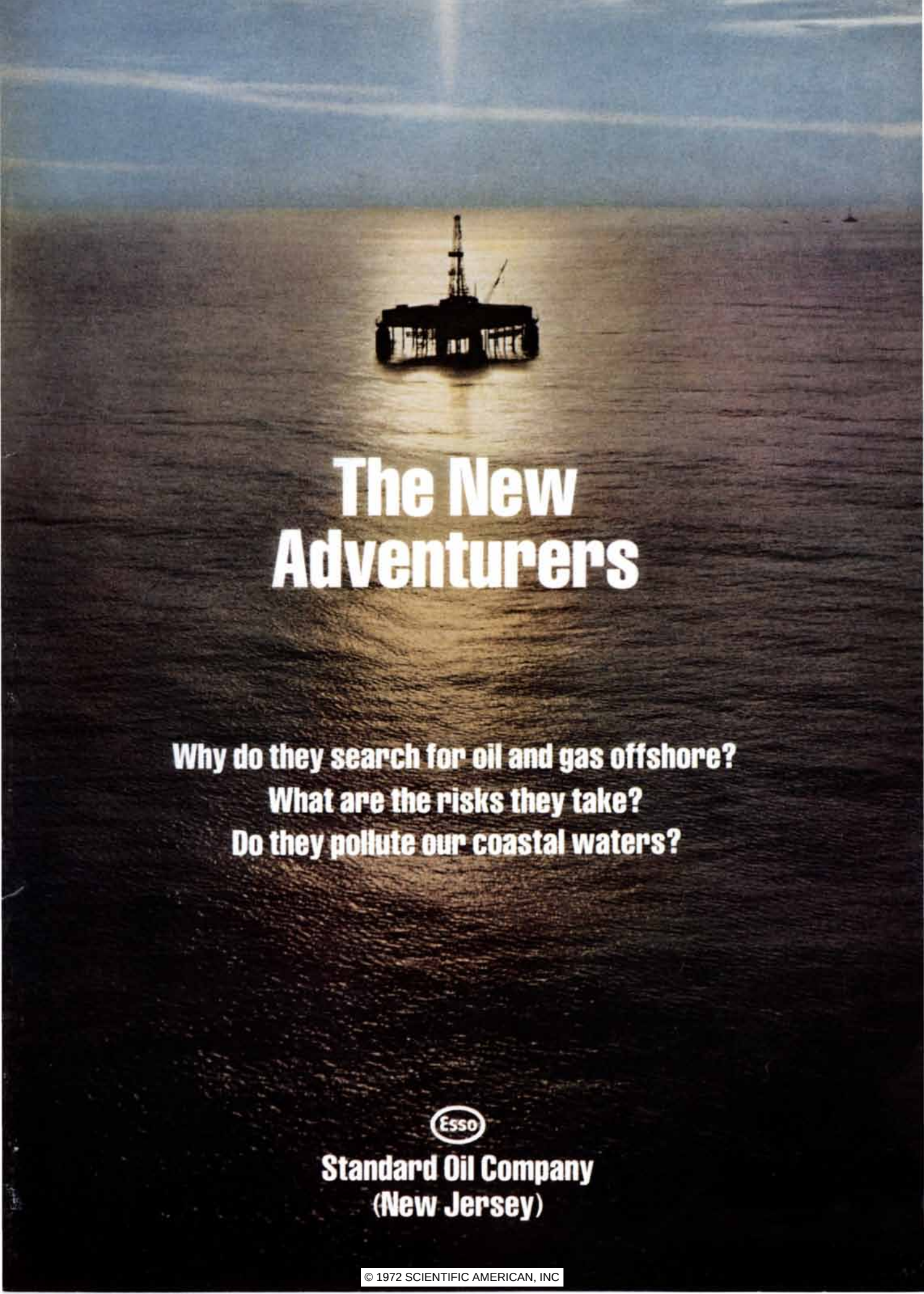
"Whereas it is very difficult, and perhaps almost impossible, to detect the adulteration of liquors, we are fortunately able to detect the adulterator of the ordinary articles of food and to detect his practices with certainty. Chicory and coffee are so unlike that the difference is easily perceived; red lead, added to vermilion, is easily separated; sulphuric acid or oil of vitriol, when used for the purpose of increasing the strength of vinegar, is readily recognized; sugar, when adulterated with sand, may easily be made to give positive evidence of the presence of the latter. Any attempts to suppress the practice of adulteration must be based on certainty of exposure and punishment. How many children are robbed of their due amount of nutriment by the watering of milk? We feel satisfied that the vile practice of adulteration will never be completely and permanently checked until the Government takes the matter fairly in hand and enacts efficient laws looking to the detection and punishment of this crime."

"The *Adriatic*, a new and splendid ocean steamer, recently made the passage from Queenstown to New York in seven days and 16 hours, the quickest westerly voyage ever made across the Atlantic."

"The pneumatic railway brake of George Westinghouse, Jr., of Pittsburgh, Pa., already in extensive use in this country, is just now being introduced abroad. In this improvement the brakes are operated by compressed air supplied from a reservoir placed under the locomotive, a special pump being employed to effect the compression. The practical results obtained in England are considered remarkable by the railway authorities there."

"A Tellier refrigerating machine, just erected in the largest brewery in New Orleans, supplies a large storeroom holding 5,000 barrels of ale and lager beer with dry cold air at a temperature of 40°, even when the temperature outside is 85°. A large refrigerating cylinder, through which pass a number of pipes, is filled with liquefied ammonia. The ammonia vaporizes, rendering the pipes through which the air passes excessively cold. The ammoniacal vapor is subsequently compressed again into liquid form and returned to the cylinder to repeat the same operation without any waste of material."

"Among the prominent men recently deceased is James Gordon Bennett, founder and proprietor of the New York *Herald*, aged 75. So far as concerns the ethics of journalism, he was unscrupulous and irregular, zealously advocating the cause of truth and justice one day and perhaps the next assailing the same cause with unworthy vehemence. The *Herald*, under Mr. Bennett's *régime*, was notoriously unreliable, but as a vehicle of news it was the embodiment of enterprise, and in this respect it outranked all its competitors of the press. The New York *Herald* is one of the most widely circulated daily papers in the world, and as a property one of the most valuable. The establishment falls, by request of its founder, to his only son, Mr. James Gordon Bennett, Jr."



(Left) Helicopters carry small equipment and men to the rigs. Crews work seven days on, then have seven days off. (Right) To replace a worn bit, the entire drill pipe is taken apart, stacked in the derrick, and then reassembled. Drillers call this oft-repeated cycle "making a roundtrip."

The New Adventurers. They search beneath the sea for new supplies of oil and gas.

A hundred miles out from the Louisiana coast over the sea, the big helicopter flies on and on. Suddenly, there appears on the far horizon a massive steel structure that seems almost delicate with its lattice of metalwork—yet with sturdy legs that fix it firmly to the floor of the ocean more than 200 feet below the waves.

This is an offshore oil well drilling platform in the Gulf of Mexico. Hundreds more like it are here in the Gulf, off the coast of California, in Alaska's Cook Inlet, the North Sea, offshore Indonesia, Africa, Australia, and in other waters around the world. Every one of them is working around the clock, seven days a week, searching for, or producing, the world's most vital sources of energy—oil and natural gas.

A career for a lifetime.

The crew is likely to be a small cross-section of America. The cook might be a Louisianan, the drillers from Texas or Montana, the geologist a New Yorker, the roughnecks from California or Oklahoma. But almost to a man, they share an interest and enthusiasm for their task, and the ability to do a job that is physically demanding. The pay is good, and the men get seven days off after every seven days on board. Offshore drilling is a special calling with a special lure. The same men return again and again to the platforms and rigs. Most of them make it a lifetime career.

In a real sense, the men who go to sea today in search of oil or gas are the "New Adventurers," the descendants of the fishermen, traders and explorers who centuries ago ventured upon the unknown for daily sustenance, wealth, or from sheer curiosity. The modern searchers on the sea also live with danger. A severe storm can force them to seek safety on shore. An oil or gas blowout can in seconds create an inferno of fire. Such disasters are infrequent but they happen—even though the most stringent safeguards are taken to avoid them.

Drilling rigs with libraries.

The living quarters on an offshore drilling platform resemble those on a modern tanker or freighter. Spotless corridors, air-conditioned staterooms and offices, a recreation room complete with a library and TV, a dining room that serves king-sized meals, where snacks and hot coffee are available at all hours. But on the working deck, it's something else: a maze of complex machinery, hoses, chains, pipes and cables. The focal point is the tall derrick, ceaselessly guiding long lengths of pipe down into the ocean floor.

A 19th century beginning.

Many people probably think of offshore drilling as a fairly recent development—especially since the 1969 blowout in the Santa Barbara Channel focused public attention so dramatically on the environmental aspects of these operations. Actually, producing oil from coastal waters dates way back to 1894. It began from piers built in that same Santa Barbara Channel. But it wasn't until 1947 that the first true offshore platform was built in fifty feet of water off the Louisiana coast. Since then, some 20,000 wells have been drilled off the coasts of 70 countries, and one-sixth of the world's oil is now

Esso

Our continental shelf. There may be more oil and gas here than has ever been found onshore in the history of the U.S. Prospects look especially good for discoveries of large volumes of urgently needed natural gas.

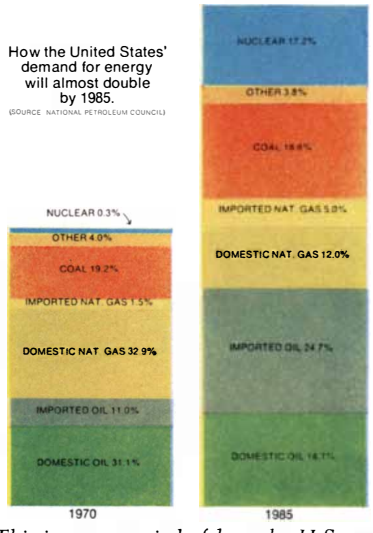
being produced from offshore fields.

Exploring for, and producing, oil and gas offshore is expensive. It costs about four times as much to drill an exploratory well offshore as a similar one on land, and operations in deeper waters will cost even more. Why then does the petroleum industry persist in its underwater search?

First of all, there's the matter of supply and demand. Beyond that are vital considerations of continued economic growth and security of supply. Right now, oil and gas provide three-quarters of America's energy requirements. Meanwhile, the demand for energy continues to increase.

Every day, the United States consumes 650 million gallons of petroleum and over 50 billion cubic feet of natural gas. This consumption is growing so fast, that the United States is expected to use as much petroleum and natural gas in the next fifteen years as it has during the entire 113 years of the oil industry's existence.

In the case of natural gas, this estimate is conservative—only because supplies are severely limited. Gas is such a clean, convenient fuel that its use would grow much faster if it were readily available. In the U.S., we are increasingly dependent on offshore areas for natural gas—a fuel which is becoming critically scarce. Overseas imports are not the entire answer, since gas is difficult and expensive to transport in anything other than pipelines.



This is one appraisal of how the U.S. may meet its future energy needs. It assumes coal users will meet air pollution regulations, construction of nuclear plants will accelerate rapidly, and availability of synthetic fuels (e.g., oil from shale) will be limited. Problem areas may be large oil imports and reduced domestic natural gas supplies.

The offshore potential.

If the U.S. is to minimize dependence on foreign energy sources, as much of our oil and gas as possible must come from domestic supplies. But it is increasingly more difficult to find oil and gas reserves on land.

That leaves the offshore areas, where the prospects for finding additional oil and gas deposits are quite encouraging. Geologists estimate the recoverable reserves of oil underlying our continental shelf may be more than the U.S. has consumed in its history. The outlook for natural gas also appears excellent, particularly along the East Coast.

The petroleum industry has turned seaward because it has had to. Today, as domestic reserves of oil and gas dwindle in relation to consumption, the offshore search is more urgent than ever. The new adventurers are embarked on a sea venture that can have far-reaching significance for our country and our people.

Esso

To look for oil and gas at sea, our explorers use modifications of the tools they use for the search on shore.

If early reconnaissance is encouraging, the search advances to seismic surveys. Shock waves are generated which travel downward, striking successive rock formations beneath the surface of the sea. These waves are reflected back to the surface where detecting devices record the impulses. By measuring the time intervals of the waves, a geophysicist can tell the general characteristics of the formations that lie below the ocean floor. But he can't tell if the rocks contain oil and gas. Only a drill can do that.

Our sleeve exploder.

A few years ago, Esso Production Research, a Jersey Standard affiliate, invented a device to replace the use of dynamite in offshore seismic work. It explodes a small gas charge in a heavy rubber sleeve, inflating it like an instant balloon and producing a seismic echo. It provides more data in less time. It's a lot less expensive than dynamite. And it won't kill the fish.

More than $800 million in offshore royalties and lease bonuses has been paid annually to state and federal treasuries.

Bids in the $millions.

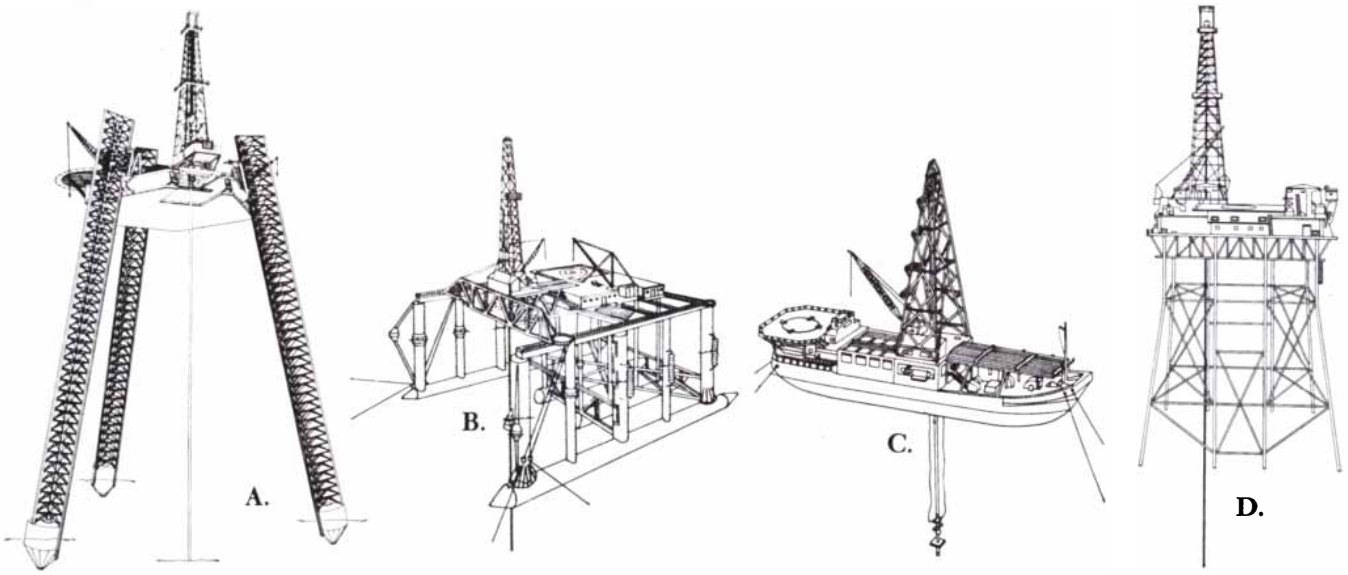
Even if our geophysical and geological findings indicate a good chance of oil or natural gas under the ocean floor, drilling is still a long way off.

Next, we have to obtain the right to drill. There may be lengthy public hearings or government studies prior to the bidding for leases.

In areas considered highly prospective, competitors for the leases can put hundreds of millions of dollars into public coffers. Oil companies must also pay royalties on every barrel of oil and every cubic foot of gas produced. In recent years, offshore lease bonuses and royalties paid to state and federal treasuries have averaged over $800 million per year.

Once a lease is granted, exploratory drilling can begin. But the odds are still against us. For every *fifty* wildcat wells drilled on land in the U.S., only *one* finds a field that will turn a profit. Because offshore areas are relatively unexplored, the odds of finding oil or gas at sea are better—but it's still a real long shot.

If exploratory drilling finds oil or gas, fixed platforms to develop a field of wells are then put in place. They contain living quarters for the work crew, a helicopter landing pad, storage space for supplies, and room for all the complicated drilling and production equipment. From these "islands," up to thirty wells can be drilled directionally to locations as far as a mile away.



In the search for oil and gas at sea, we use three basic types of mobile drilling rigs. A) This self-elevating rig is towed to a site where its legs are lowered to the sea bed, and its platform "jacked up" above the water. Jack-up rigs are limited to waters up to 300 feet deep. Beyond this depth, floating rigs must be used. B) Semi-submersible rigs are supported by gigantic buoyant chambers, which are ballasted with sea water and submerged. The drilling platform stays safely above the waves' reach. C) In very deep waters, ships with large holes midship for drilling are used. D) After a discovery, fixed platforms are built to drill development wells and handle production.

Esso

1. Drilling rig.
This semi-submersible rig cost about $30 million to build, and is leased to us on a daily basis. The upper deck is 200 feet square, and from the bottom of the vessel to the top of the rig is almost 300 feet.

2. Anchors.
Eight mooring lines extend as far as one mile from the rig. At the end of each is a 30,000-lb. anchor.

3. Drilling pipe and riser.
The drill pipe, a string of 30-foot pipe lengths with the drilling bit attached, runs through the riser. The riser, a 2-foot-diameter casing extending from the rig to the wellhead, guides the drill pipe into the hole. Together, they also circulate the specially compounded drilling fluid called "mud."

4. Underwater eyes.
A television camera is lowered to the ocean floor to check equipment. It sends back pictures of every joint, every bolt, every valve within its sight.

5. Blowout preventer.
It is attached firmly to the wellhead on the ocean floor when drilling starts. This one, built specifically for use in deep waters, is 30 feet high and weighs 40 tons. It is operated by hydraulic lines running to the surface. If unexpected pressures are encountered, it is designed to control them until normal drilling can be resumed.

6. Underground casing.
As the bit drills deeper, the hole is periodically lined with steel and cement. This prevents caving and seals off high pressures.

7. Drilling bit.
In hard rock, a bit may drill only a few feet before it gets dull. In soft shale, it can cut through 100 feet an hour, and last a day or more.

8. Core sample.
Oil and natural gas are produced from sedimentary rocks, like this 4-inch-diameter sandstone core cut by a special bit. Underground "pools" of oil exist only in the imagination.

Today, hundreds of producing platforms rise from the Gulf of Mexico. Fish thrive in great abundance and variety among the platform legs, which act as artificial reefs. Both sport and commercial fishermen agree the fishing is better than ever.

"No Pollution—Not Any." This order is posted throughout our drilling rigs.

The design and construction of offshore structures is an essential part of the petroleum industry's efforts to protect the ocean environment. To insure that platforms can withstand the elements, extensive research has been done on the effects of wind and wave forces that develop at sea, on water depths, currents, and sea floor conditions. By reducing the risk of damage to these installations, the possibility of pollution-causing accidents is reduced.

Built to last.

Structures are designed for local conditions. In the Gulf of Mexico, platforms have endured full hurricane winds, and in the North Sea, the violent storms that whip down from the Arctic. A special challenge is withstood by platforms in the Cook Inlet of Alaska: ice floes four feet thick. During the winter these are carried in and out of the inlet by swift currents and thirty-foot tides.

You might suppose that offshore operations would have a harmful effect on traditional fishing grounds. Just the opposite has happened. In a few months, a new platform becomes an artificial reef with tiny marine life clinging to the underwater structures. Small fish feed on these and, in turn, attract larger fish, which attract sport fishermen. And drilling doesn't seem to have harmed commercial fishing. In fact, Gulf Coast fleets now haul in four times the weight of fish they caught before the drilling started.

Drilling an offshore well is an around-the-clock operation—one that must be closely supervised at all times. Many mechanical and electronic devices log the progress of the drilling, and keep a continuous check on conditions in the well.

"Mud" is more than mud.

During the drilling process a special fluid called "mud" (a mixture of water, clay, and mineral additives) is pumped down the drill pipe under carefully controlled pressure. When the "mud" reaches the bottom of the hole, it is forced out through the drilling bit and returns to the surface in the space between the drill pipe and the walls of the hole.

This constantly circulating fluid has several functions. It cools and lubricates the drill bit. It flushes from the well the cuttings left by the bit. It cakes and stabilizes the sides of the hole, preventing cave-ins. And it acts as a safety device, chiefly by its weight, to contain underground pressures. High-pressure zones are counteracted by increasing the weight of the "mud."

As a standard pollution-control practice, any drilling "mud" which could damage the environment is sent ashore for disposal. The same applies to test samples such as cores and fluids —and to all wastes generated by the drilling crews. One of our rules—strictly enforced—is that not so much as a paper cup should go over the side.

Like mud, "casing" serves multiple roles in drilling operations. Casing consists of long lengths of steel pipe which are set in the drill hole. It helps maintain good circulation of the mud, prevent erosion of the walls of the hole, protect fresh water strata, and shut off high-pressure zones. Once casing is set in the hole, it is sealed securely in place. This is done by pumping a cement slurry, or mixture, down the inside of the casing. It is forced out through the bottom, and rises to fill the space between the outside of the casing and the walls of the drill hole.

After the first "string" of casing is

Esso

This is a sixteen-foot model of what someday may be the tallest offshore platform. It will be nearly as tall as the Empire State Building, but only the tip will rise above water. The model was used to determine how such a huge structure can be safely towed and submerged in place.

In deeper water, we'll need platforms as tall as skyscrapers.

set, "blowout preventers" are installed on the well. These highly important safety devices can close off the space between the drill pipe and the casing in a matter of seconds if unexpectedly high pressures are met, and contain these pressures until normal drilling can be resumed. When the well is ready for production, a system of high-pressure surface valves, called a Christmas tree, is installed on top to control the well flow. Or, if the well does not flow naturally, a pump can be added. On the platform, the oil and gas are separated and sent to shore through pipelines.

Onshore and offshore, between thirty and forty thousand oil and gas wells are drilled around the world every year. Yet, only a handful result in those violent eruptions called "blowouts"—but the possibility is still there.

A prevention paradox.

Prevention is the most effective way to deal with the problem of blowouts. To that end, Jersey has established training facilities in the U.S. and abroad for drilling superintendents and contract crews from our affiliates all over the world.

Our main school is centered around a mile-deep well on the King Ranch in Texas. This practice well is the real thing. It has storage tanks, a pump, control manifold, a full array of pressure gauges and controls—all the diverse equipment that is used in the field to prevent a blowout.

Oddly, all this training represents something of a paradox, set up mainly because there are so few blowouts. Since drillers may never experience a runaway well, the schools give them an opportunity to improve their skills in control and prevention.

Four times deeper.

As the demand for energy grows steadily, the search for new sources of oil and gas has moved ever farther offshore and into deeper waters. This has necessitated refinement and extension of the methods used for many years in shallower waters.

Recently, Humble has devised a system that will make possible the development and production of oil and gas in waters ranging from 700 to 1500 feet deep. This is from two to four times as deep as any offshore fields now producing.

A basic component in the Humble deep-water system will be, of course, drilling platforms. The first one will be a real giant. Plans call for a huge four-legged structure in nearly 800 feet of water, and weighing more than 20,000 tons. As many as sixty wells can be drilled directionally from its 150-foot-square deck. Each of the giant's four legs will be seventeen feet in diameter, and they will straddle more than an acre of the ocean floor.

The SPS.

The second part of the project—for the deepest waters—will be Submerged Production Systems, secured to the bottom of the sea. Each SPS will include a cluster of up to forty directional wells drilled from floating rigs.

The whole underwater operation involves advanced technology, with electronics to monitor and "instruct" each unit, hydraulics to open and close valves, and safety devices that will automatically close off any part of the system that malfunctions.

Oil and gas from each SPS will flow through pipelines to gathering points on centrally located platforms, and from there to shore.

Esso

The Santa Barbara Channel today. In 1969, a blowout caused a terrible, but temporary, blight. Since the clean-up, scientists have found no evidence of lasting damage to sea growth, marine life or the beautiful beaches.

The tragic Santa Barbara blowout wasn't at one of our wells. But that's beside the point; we do drill there. That blowout caused the death of about four thousand birds, and a terrible, but temporary, blight to a beautiful shoreline. The entire oil industry is paying for this accident—in millions of dollars' worth of drilling delays, and a lot more in reputation.

Our industry's excellent safety record has been marred severely by that one event.

We aren't trying to make excuses for what occurred, but some of the after-effects seem to have been exaggerated. A research team from the University of Southern California confirmed there was no apparent lasting damage to sea growth, marine life, or beaches as a result of the spill.

Some lasting benefits.

To a lot of people, the blowout ushered in the age of ecology. It also made the oil industry and the Government take a good hard look at operating procedures. And some lasting benefits are evolving.

Both private industry and the U.S. Coast Guard are now working on a large variety of oil containment and recovery systems to minimize the damage caused by blowouts or spills. Where containment and recovery are impractical, research has developed nontoxic dispersants and biodegrading agents to assist in cleaning up.

The Santa Barbara blowout taught us all a painful lesson.

Also, extensive research is being done on keeping our beaches clean. Surprisingly, most authorities agree ordinary, plentiful straw is still one of the most effective ways to absorb oil.

Clean Seas Inc.

One significant development has been the creation of Clean Seas Inc. It was set up and financed by Humble and 14 other oil companies operating in the Santa Barbara Channel. Its purpose is to provide immediate equipment, materials, manpower and know-how to handle spills of all types. Clean Seas is just one of 70 organizations around the country set up to meet such emergencies in coastal areas.

We work with people and machinery and there is always the possibility of an accident. Yet in completing over 1500 wells in U.S. waters over the years, Humble has had only three blowouts. All were controlled without polluting the water or the shore line.

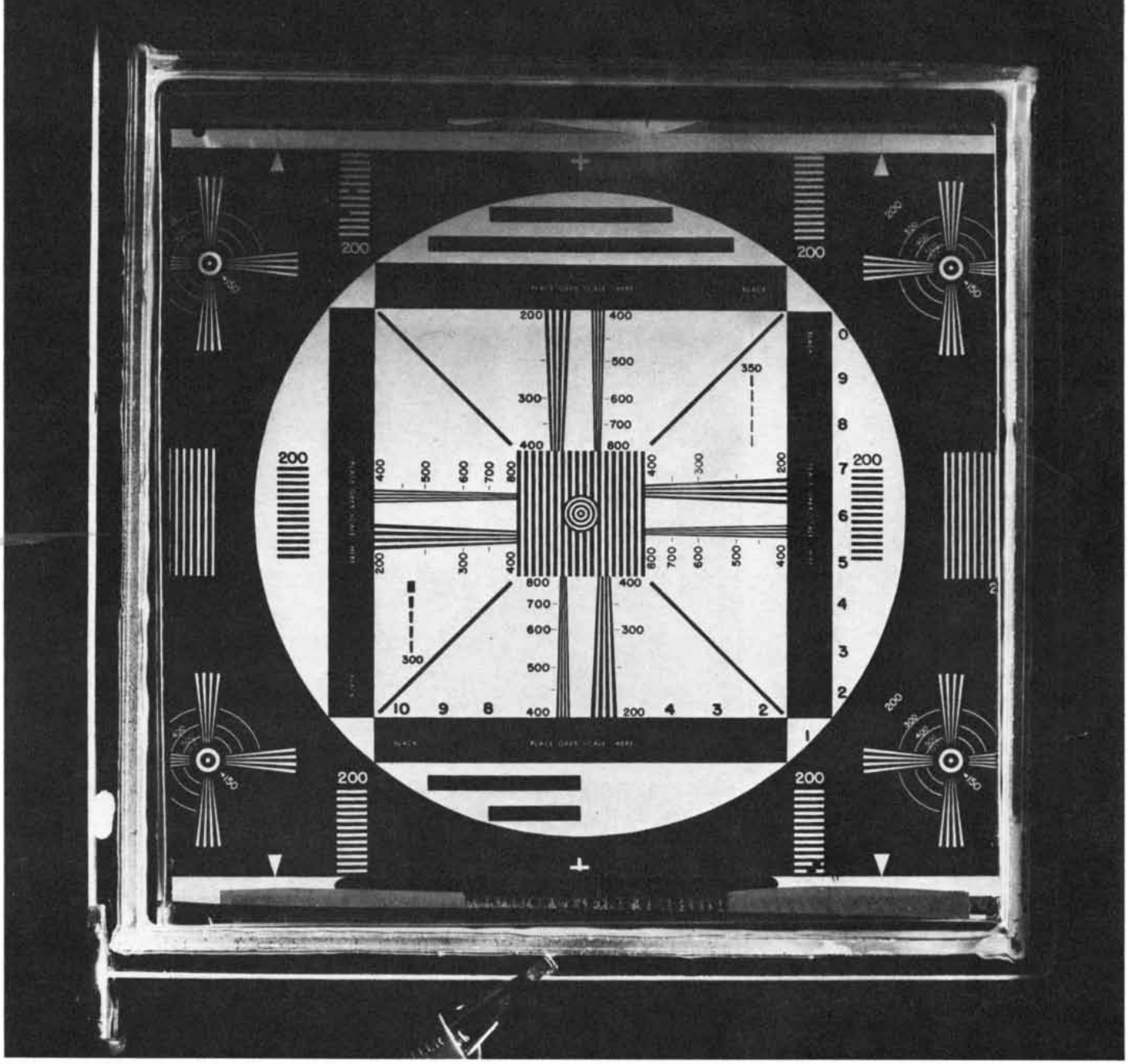
We can't promise we'll never have a damaging accident. But we can promise that as long as we work offshore we'll operate in the most responsible way we know how.

A last word.

The United States is no longer in an era of abundant energy. Our needs for the future are critical. But not everyone agrees where the energy should come from. Decisions for or against offshore development, or any other source of energy, are too important to be made by less than an informed public. In these pages, we have presented the case for offshore development as we see it. If you would like additional copies of this report, write to "The New Adventurers," Standard Oil Company (New Jersey), 1251 Avenue of the Americas, New York, N.Y. 10020.

Esso

Standard Oil Company (New Jersey)



A LIQUID-CRYSTAL SCREEN DISPLAYS TV TEST PATTERN

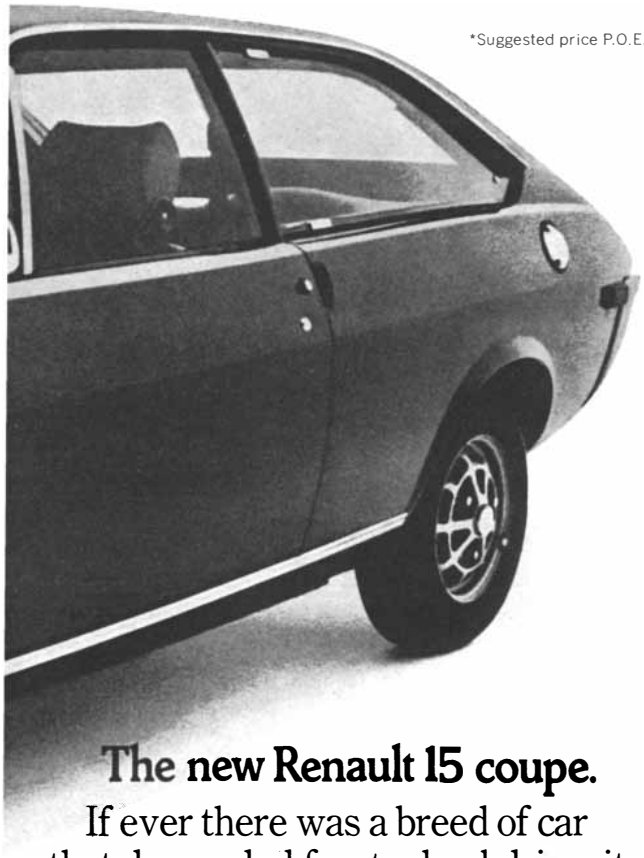
WILLIAM VANDIVERT

THE SEPTEMBER ISSUE OF **SCIENTIFIC AMERICAN** WILL BE DEVOTED TO

COMMUNICATION

Presenting an obvious but brilliant idea.

Front-wheel drive in a sporty coupe.



*Suggested price P.O.E. Slightly higher on the West Coast. Freight, taxes, options and preparation additional.
Renault, Inc., 100 Sylvan Avenue, Englewood Cliffs, N.J. 07632

The new Renault 15 coupe.

If ever there was a breed of car that demanded front-wheel drive, it is the sporty coupe.

Until now, a lot of sexy 4 seaters that look like dream machines in the driveway have embarrassed their owners on the road. In designing our sporty car, we did the obvious. We gave it front-wheel drive to make sure its handling lived up to its looks.

Instead of being shoved from behind by fixed wheels, you are pulled along by the same wheels you're steering with for better control. And because the engine's weight is over them, those wheels really take a good bite into the road.

You have, in other words, a terrific looking coupe that actually holds the road a lot better than many so-called sports cars.

What the front-wheel drive pulls around is also noteworthy: It has a tough aluminum engine, double-barrel carburetor, ultra precise rack-and-pinion steering, steel-belted radial tires, front disc brakes, stabilizer bars front and rear, a smooth 4-on-the-floor synchromesh gear box, a roomy interior with 4 sinfully plush seats and a price of $3325.*

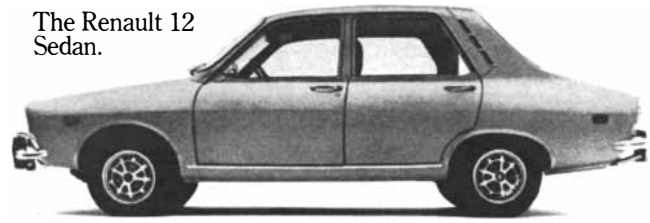
Above all, with front-wheel drive, it is quite simply a better-handling sporty coupe. Period.

In fact, while they aren't as roguish in appearance, our other Renaults have much in common with the Renault 15. Including the uncanny road holding of front-wheel drive.

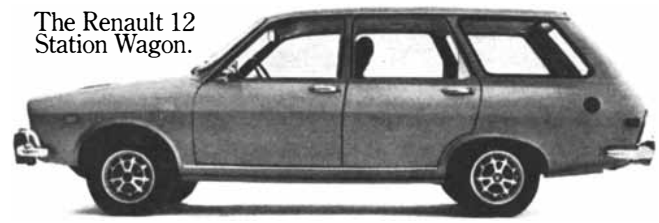
Which is just as good an idea for non-sporty cars.

But not as obvious.

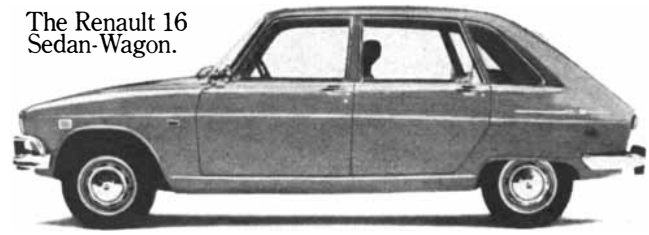
The Renault 12 Sedan.



The Renault 12 Station Wagon.



The Renault 16 Sedan-Wagon.



RENAULT

World's largest producer of front-wheel drive cars.

THE AUTHORS

HERBERT SCOVILLE, JR. ("Missile Submarines and National Security"), is secretary of the Federation of American Scientists and chairman of its strategic weapons committee. Following his graduation from Yale University in 1937 he did graduate work in physical chemistry at the University of Cambridge and the University of Rochester, obtaining his Ph.D. from Rochester in 1942. He has served as deputy director for research of the Central Intelligence Agency and assistant director for science and technology of the U.S. Arms Control and Disarmament Agency and also as director of the arms-control program of the Carnegie Endowment for International Peace.

RALPH A. REISFELD and BARRY D. KAHAN ("Markers of Biological Individuality") are respectively a senior member of the department of experimental pathology at the Scripps Clinic and Research Foundation in La Jolla, Calif., and surgical resident at the Massachusetts General Hospital of Harvard University. In July they will join each other to develop an integrated transplantation and tumor immunology clinical and research unit at Northwestern University. Reisfeld, who was born in Germany, received his bachelor's degree at Rutgers University in 1952 and his Ph.D. from Ohio State University in 1957. He spent four years at Merck & Company and then worked at the National Institute of Allergy and Infectious Diseases before going to La Jolla in 1970. He describes his hobbies as reading "science fiction and biographies of interesting political figures," going on fishing trips and making excursions to the nearby mountains with his family. Kahan holds both a Ph.D. (in physiology) and an M.D. from the University of Chicago, having obtained them in 1964 and 1965 respectively. Starting at the Massachusetts General Hospital in 1965 as an intern in surgery, he has been there since except for two years (1966 to 1968) as a staff associate in the laboratory of immunology of the National Institute of Allergy and Infectious Diseases. His outside interests include playing piano duets with his wife and conducting pre-Columbian archaeological explorations.

JAMES G. LAWLESS, CLAIR E. FOLSOME and KEITH A. KVENVOLDEN ("Organic Matter in Meteorites") are members of a team doing research on meteorites at the Ames Research Center of the National Aeronautics and Space Administration. Lawless, who is a mass spectroscopist, obtained his Ph.D. from Kansas State University in 1969 after studies at Purdue University and Lafayette College. Folsome, whose Ph.D. is from Harvard University, is professor of microbiology at the University of Hawaii. Kvenvolden is an organic geochemist, currently head of the chemical-evolution branch at Ames, and a consulting associate professor at Stanford University. He was trained at the Colorado School of Mines and holds a Ph.D. from Stanford.

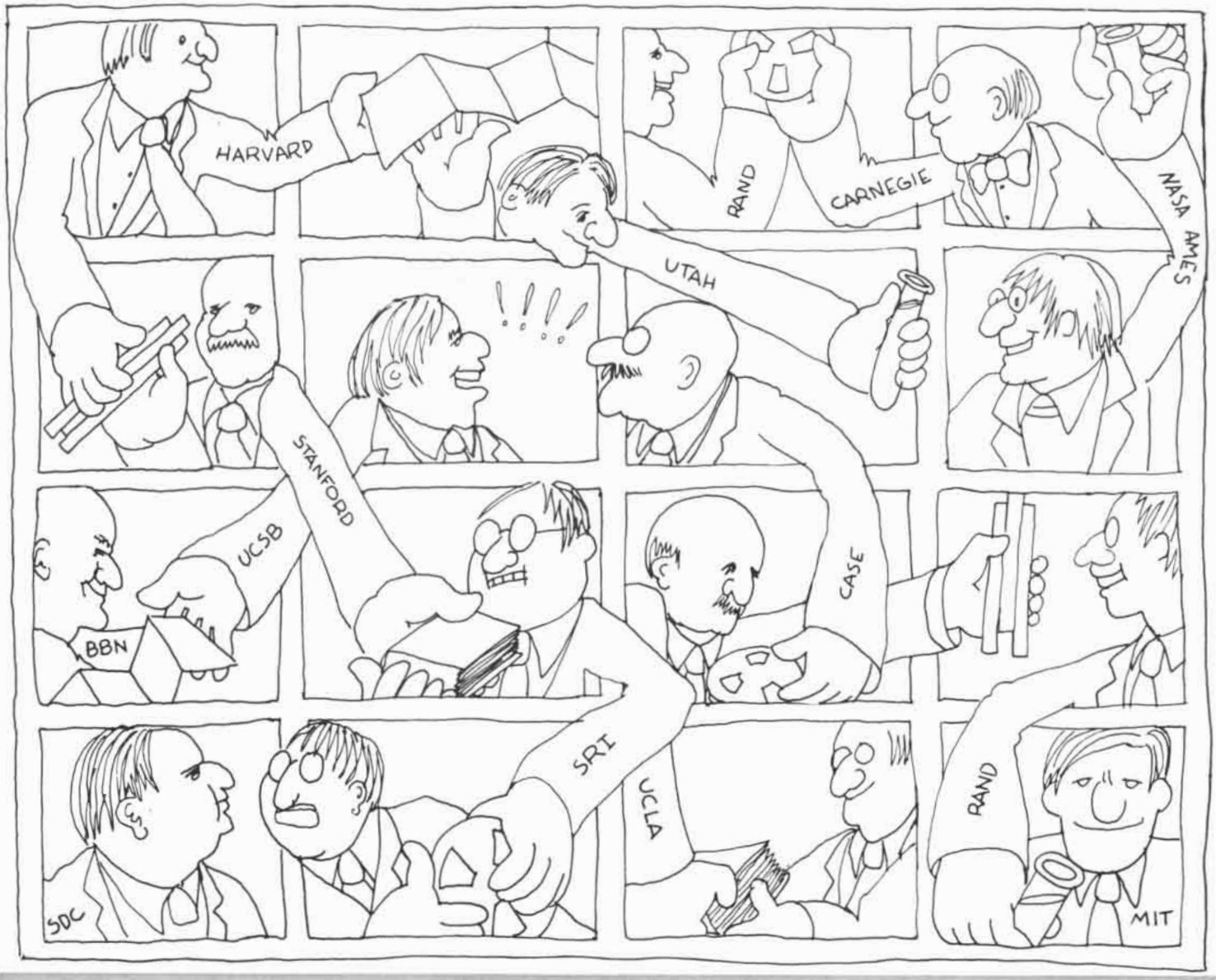
NORMAN D. NEWELL ("The Evolution of Reefs") is a geobiologist at the American Museum of Natural History and Columbia University. He obtained his Ph.D. from Yale University in 1933 after earlier training at the University of Kansas. Of his professional and private interests he writes: "I champion the adequacy and richness of the fossil record of life, which I believe is hardly sampled. I am particularly interested in the great episodes of extinction of animals and plants in geologic history and the sometimes delayed rebound of evolutionary activity."

BERND HEINRICH and GEORGE A. BARTHOLOMEW ("Temperature Control in Flying Moths") are respectively assistant professor of entomology at the University of California at Berkeley and professor of zoology at the University of California at Los Angeles. Heinrich writes: "I grew up on a farm in Maine, where our family settled after coming to this country from Germany after the war. In Maine I spent as much time as possible in the woods, lining bees to find wild-bee trees and hunting deer in the fall, snowshoeing and hunting rabbits in the winter and fishing and canoeing in the spring and summer. At the University of Maine I began in forestry and graduated in zoology after taking 14 months off to participate in an ornithological expedition to East Africa. To earn college expenses I cruised timber in the northern part of the state and worked as a laboratory assistant." Heinrich obtained his Ph.D. at U.C.L.A. in 1970. Having been a track and cross-country man in college, he continues to run; he is also interested in photography and biological illustration. Bartholomew was graduated from the University of California at Berkeley in 1940 and obtained his master's degree there in 1941; his Ph.D., obtained in 1947, is from Harvard University. He has been at U.C.L.A. since 1947, working (as he puts it) "at the interface between physiology, behavior and ecology."

MARTIN DAVIS and REUBEN HERSH ("Nonstandard Analysis") are respectively professor of mathematics at the Courant Institute of Mathematical Sciences of New York University and professor of mathematics at the University of New Mexico. Davis writes: "My parents were hard-working Polish-Jewish immigrants. Most of my childhood was spent in a slum neighborhood in the East Bronx. I attended the Bronx High School of Science (graduated 1944) and City College (B.S. 1948)." Davis took his Ph.D. at Princeton University. "I count myself a political radical," he says, "although in trying to align myself with my fellow radicals, I am usually disconcerted by a choice between dogmatism and mindless anti-intellectualism." Hersh was graduated from Harvard College with a degree in English literature. After four years as an office boy and editorial assistant with SCIENTIFIC AMERICAN and four years as a machinist he obtained his Ph.D. in mathematics from N.Y.U.

FLOYD RATLIFF ("Contour and Contrast") is professor of physiological psychology at Rockefeller University. He was graduated from Colorado College in 1947 after five years of service in the U.S. Army as an artilleryman. After receiving his doctor's degree from Brown University in 1950 he was in the psychology department at Harvard University for three years before moving to Rockefeller University (then the Rockefeller Institute) in 1954. "I am much interested in the history and philosophy of science," he writes, "mainly in the contributions of the physicist-philosopher-psychologist Ernst Mach, and in the relations between art and science."

JAY M. WEISS ("Psychological Factors in Stress and Disease") is in the laboratory of physiological psychology of Rockefeller University. "After spending most of my youth on basketball courts, tennis courts and ball fields in Passaic, N.J., and even some time in bowling alleys and poolrooms," he writes, "I went on to Lafayette College for an A.B. degree and from there to Yale University for a Ph.D. in 1967. I am particularly interested in applications of experimental biology to abnormal psychology. I still play as much tennis as I can get in; I often think that if I could hit a decent backhand, I might throw over research and take up tennis seriously."



ARPA has a network of Supercomputers.

There are two dozen huge computer systems in the Advanced Research Projects Agency network.

Over half of them are DECsystem-10s. Our Supercomputer. MIT has two. So does Utah. Then there's Harvard, BBN, Carnegie, Case, SRI, Stanford and Rand.

Which should give you some idea of how popular our DECsystem-10 really is.

In the ARPA network, DECsystem-10's are doing state-of-the-art research into weather forecasting, econometric studies, resource management, computer sciences, and much more. Everyone shares their computer and expertise with everyone else. Everyone comes out ahead.

Additional DECsystem-10's will be a part of ARPA's mammoth ILLIAC IV number crunching complex now being installed at NASA-Ames Research Center in California. They'll handle all communications while at the same time managing up to a trillion bits of file storage. Once ILLIAC IV is in gear, ARPA members will be able to do in hours jobs so big that they wouldn't even attempt to do them now.

ARPA is one of the biggest brain trusts ever assembled. If half of its members have a DECsystem-10, you really ought to know about it.

Write for the literature that explains why 62 of the leading universities and research institutions in the country have selected DECsystem-10's. (Hint: It does computation *and* timesharing at half the cost of other systems — without sacrificing throughput.)

DECsystem-10 Scientific Group. Digital Equipment Corporation, 146 Main St., Maynard, Mass. 01754. (617) 897-5111.

digital

A CHIEF EXECUTIVE OFFICER SPEAKS HIS MIND. PORTION OF AN INTERVIEW WITH
JOHN O. LOGAN, PRESIDENT, UNIVERSAL OIL PRODUCTS COMPANY, DES PLAINES, ILLINOIS

Q: Is it true, Mr. Logan, that the nation is running out of gas?

A: Yes. Our proven natural gas supply will be exhausted by the 1990's.

Q: What can be done about it – now?

A: We at UOP are into three different solutions operationally—as a result of this prediction by the Federal Power Commission.

We are just completing construction of the largest liquified natural gas plant in the world on the North coast of Borneo. We process at extremely low temperatures; then cryogenic tankers transport and release the liquid gas directly into pipelines, anywhere.

Second, we are engineering the largest substitute natural gas plant in the world at Plainfield, N. J. A methane-rich substitute "natural gas."

Third, we are into coal "gasification"—directly substitutable "natural gas" from coal. America's coal supply is good for another 200 to 300 years!

All these ventures are possible because UOP is an integrated-systems company that can take an idea to commercialization rapidly and within the corporate family. The know-how, the people and the facilities to resolve a need from concept to final construction. It's an extremely rewarding and satisfying way to run a business.

Write to Office of The President for details on:
( ) UOP's stake in the Energy Crisis
( ) UOP's stake in the Environment Crisis
( ) UOP's total Capabilities story

UOP, Universal Oil
Products Company
Ten UOP Plaza
Des Plaines, Illinois 60016

uop
The Answer Company

Established 1845 SCIENTIFIC AMERICAN June 1972 Volume 226 Number 6

Missile Submarines and National Security

Land-based missiles are giving way to submarine missiles as a secure deterrent to a nuclear first strike. The question now is whether or not the U.S. should spend perhaps $40 billion on a new missile fleet

by Herbert Scoville, Jr.

In recent years, as the nuclear-weapons arsenals of both the U.S. and the U.S.S.R. have continued to grow, the concept of deterrence has become almost universally accepted as the key to maintaining national security and preventing the outbreak of a nuclear war. "Winning" a nuclear exchange is no longer regarded as a rational strategic objective; in such an exchange everyone, participant and nonparticipant alike, would be a loser. In keeping with the deterrence principle President Nixon affirmed in his State of the World Message of February 9 that "our forces must be maintained at a level sufficient to make it clear that even an all-out surprise attack on the U.S. by the U.S.S.R. would not cripple our capability to retaliate." For the Russians to feel secure they must have a similar capability; only then would a stable strategic balance exist.

The primary attribute required of any deterrent force is the ability to survive a "counterforce," or preemptive, attack. Ballistic-missile submarines are almost ideally suited to satisfying this requirement. Although they are expensive compared with other strategic weapons (more than $100 million per submarine exclusive of the missiles), their mobility and invisibility make them virtually immune to destruction in a surprise attack. In contrast, land-based intercontinental ballistic missiles (ICBM's) can readily be located with the aid of surveillance satellites, so that they must be regarded as "targetable" in the event of an enemy first strike. Attempts to "harden" such fixed missile-launchers (that is, to increase their resistance to the effects of nuclear explosions) are in the long run doomed to futility, since in the absence of qualitative arms-control agreements improvements in offensive missiles, particularly improvements in accuracy, will inevitably make fixed missile-launchers vulnerable and hence reduce confidence in their deterrence value.

The advent of multiple independently targetable reentry vehicles (MIRV's),

	SUBMARINES	MISSILES	WARHEADS		MAXIMUM TARGETS	
			BEFORE MIRV	AFTER MIRV	BEFORE MIRV	AFTER MIRV
U.S.	41	656	1,712	5,440	656	5,120
U.S.S.R.	26 (42)	416 (672)	416 (672)		416 (672)	

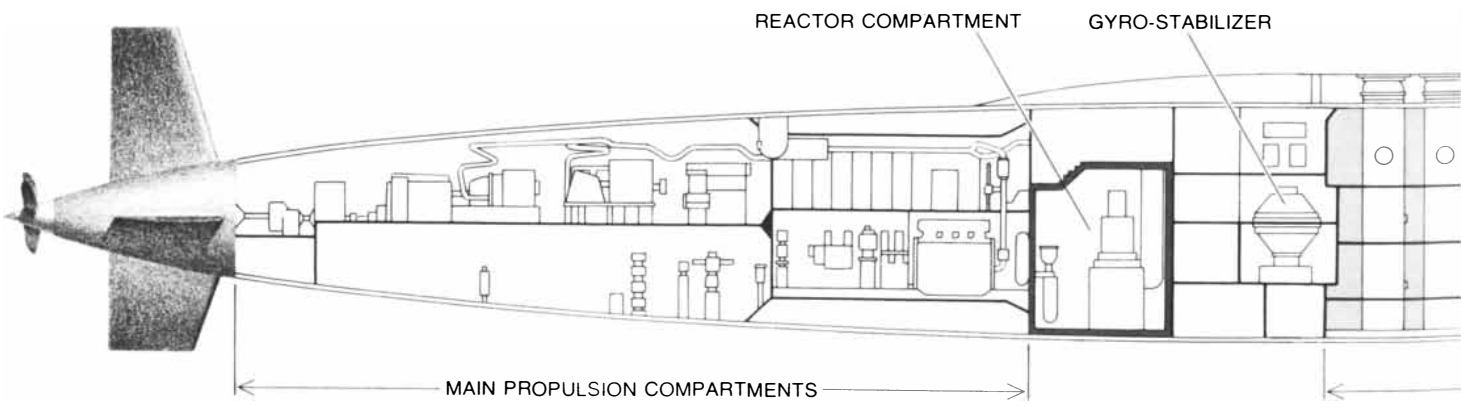
MISSILE-SUBMARINE FORCES of the U.S. and the U.S.S.R. are compared in this table. The submarines of both fleets are designed to carry 16 ballistic missiles each. The U.S. is currently in the process of converting 31 of its total of 41 deployed Polaris submarines to carry Poseidon missiles, each of which is capable of carrying up to 14 multiple independently targetable reentry vehicles (MIRV's); on the average each Poseidon will be able to deliver 10 nuclear warheads on 10 separate targets or at intervals on the same target. (The remaining payload can be devoted to various "penetration aids" intended to foil a potential enemy anti-ballistic-missile, or ABM, system.) Before the MIRV program began the bulk of the U.S. missile-submarine fleet was armed with Polaris *A-3* missiles, which feature a multiple reentry vehicle (MRV) capable of carrying three warheads; these warheads are not, however, widely separable and are aimable only in shotgun fashion at a single target. After MIRVing is complete the remaining 10 ships in the Polaris fleet will continue to be armed with *A-3* missiles. The Russian missile-submarine fleet consists at present of 26 ships deployed and about 16 more under construction, but this program could continue. (Known future submarine and missile totals are given in parentheses.) Since its missiles carry only one warhead each, however, the number of warheads it can launch is no greater than the total number of its missiles. Moreover, in a crisis military strategists in the U.S.S.R. must add to the U.S. totals the missile-submarine forces of France and Great Britain, each of which may eventually consist of four submarines and 64 missiles.

which are currently being deployed on a large scale by the U.S., creates a situation in which the "exchange ratio" strongly favors the attacker. Thus a single missile with, say, six warheads can potentially destroy six enemy ICBM's if they are caught in their silos. Moreover, strategic bombers are extremely vulnerable while they are on the ground and would therefore be very susceptible to annihilation in a surprise missile attack. Attempts to avoid this weakness by maintaining aircraft on continuous airborne alert have proved to be expensive and potentially dangerous. Even the current 15-minute ground alert is not completely satisfactory, since adequate warning would be more difficult to obtain if fractional-orbital-bombardment systems (FOBS) or depressed-trajectory missiles launched from submarines were used to attack the bombers.

Hence given the present state of military technology and reasonable anticipated advances, the primary element in the strategic-deterrent forces of both the U.S. and the U.S.S.R. will continue to be the ballistic-missile submarine. All other strategic systems will remain secondary. Moreover, it seems likely that any agreement that may emerge in the near future from the strategic-arms-limitation talks (SALT) will further enhance the relative importance of the missile-submarine forces. Since the chances that MIRV's will be limited by a SALT agreement are extremely low, ICBM's will become increasingly vulnerable. The more likely limitations on anti-ballistic-missile (ABM) systems, on the other hand, would guarantee the retaliatory capability of even a comparatively small number of submarine-launched ballistic missiles (SLBM's). The expected failure to limit antiaircraft defenses and to restrain qualitative improvements in offensive-missile systems would further decrease the value of strategic bombers. Although there will probably not soon be restrictions on antisubmarine-warfare (ASW) measures, the technology in this area is so far behind that it could not possibly threaten the submarine deterrent, if it can threaten it at all, until far in the future. In sum, the Navy will increasingly be the principal military guardian of our national security.

What characteristics must an SLBM force have in order to fulfill its function as a deterrent against the initiation of nuclear warfare by the U.S.S.R.? (Since China is so far behind both the U.S. and the U.S.S.R. in this respect, the same forces would be more than adequate to deter China as well.) First of all, the submarines should be designed to operate in, and fire their missiles from, large enough ocean areas in a variety of directions around the U.S.S.R. so as to decrease their vulnerability to ASW detection and tracking and to facilitate the penetration of any ABM system. The closer these areas are to ports in the U.S., the less will be the time lost in moving to and from operational stations and the less will be the need for overseas bases. Higher submarine speeds will also reduce this travel time and increase the ability to break contact with a trailing ASW submarine or surface vessel. The gains here may be marginal, particularly since tracking vessels will probably be faster than any missile submarine. The faster a submarine moves through the water, however, the more noise it will produce, and in countering ASW measures quietness may be much more important than speed. The reduction of submarine noise is the most critical element in preventing detection and continuous covert tracking, both of which must rely on passive acoustic sensors.

If an ABM defense is a realistic possibility, then the submarine missiles must have enough payload capacity to allow the use of multiple warheads and other penetration aids. The entire submarine force should be large enough so that the destruction of a few submarines by a concerted enemy attack, by slow attrition or perhaps by a series of accidents does not seriously degrade its overall capability. If continuous tracking by antisubmarine submarines or other ASW



TYPICAL POLARIS SUBMARINE, in this case an advanced model belonging to the Lafayette, or 616, class is shown schematically in the cutaway drawings on these two pages. The nuclear-powered submarine has a length of 425 feet, a beam of 33 feet and a submerged displacement of 8,250 tons. In addition to its 16 ballistic-missile silos it is equipped with four bow torpedo tubes. Each of its

vessels ever becomes a realistic operation on a large scale, then the more vessels there are in the missile-submarine fleet, the harder it will be for this tactic to be successful in destroying the entire force. Ballistic-missile submarines cannot be used to attack other submarines and are no threat to the SLBM deterrent of the other side.

In addition to an adequate number of submarines, missiles and warheads, it is essential to have secure and reliable communications between these vessels and their command authorities. It is not enough to send the submarines to sea with sealed orders. Controls to prevent inadvertent or unauthorized firing are an absolute necessity, and reliable methods for ordering retaliation in the event of a surprise attack are required. These communications must be jam-proof; the potential attacker cannot be allowed to hope that a communications failure might prevent a retaliatory strike.

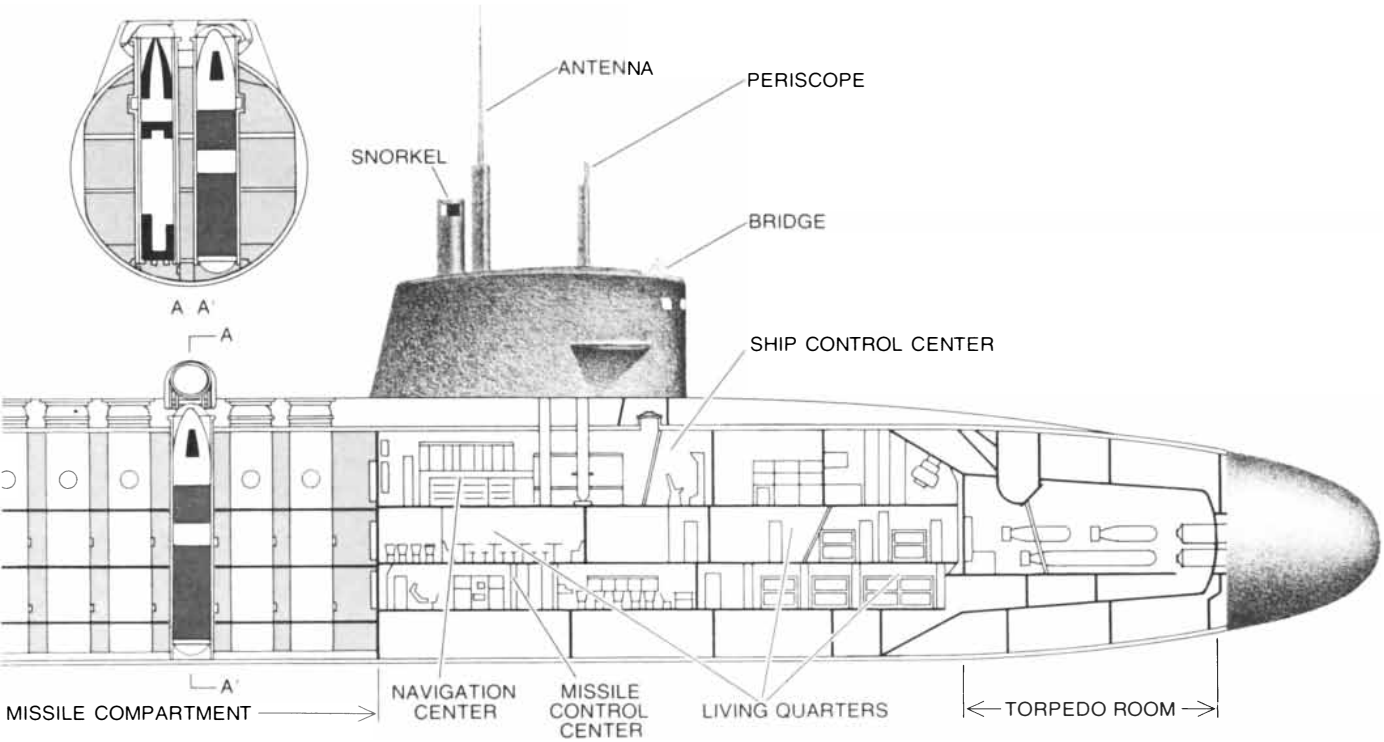
The submarines must also be able to navigate accurately, so that after they have moved through the oceans into their operational areas they will always be in a position to fire their missiles at predetermined targets. High navigational accuracy is not as great a requirement for a retaliatory strike against cities as it would be if the submarines were to be used in a counterforce role for destroying such "hard" targets as enemy missile sites. In fact, if one side wishes to use missile submarines only as deterrent weapons, then it is important that the accuracy-yield combination of the system not be so great as to give the other side concern that the submarines have a first-strike capability against land-based ICBM's; otherwise the position of mutual stable deterrence will be eroded.

With these general principles in mind, let us examine how the U.S. missile-submarine forces have developed over the years. The U.S. launched the first nuclear-powered submarine, the *Nautilus,* in 1955, but it was not until the late 1950's that development of long-range missiles had proceeded to the point that these could be installed in such submarines. The first ballistic-missile submarine, the *George Washington,* became operational in November, 1960. It was armed with 16 solid-fuel Polaris *A-1* missiles, which could be fired at a rate of about one per minute. The range of this missile was about 1,200 nautical miles and the warhead yield about one megaton. The submarines were designed to fire their missiles while submerged, using compressed gases to expel the missile; the rocket engine is then ignited after the missile has cleared the surface. By 1963, 12 more Polaris submarines were operational.

Meanwhile the development of more advanced missiles continued. The next-generation missile, the *A-2,* had a range of about 1,500 nautical miles. The first test of the *A-3* missile, with a range of 2,500 nautical miles and a "triplet" reentry vehicle, was conducted from a submerged submarine in the fall of 1963. The triplet reentry vehicles, which could carry three individual nuclear warheads each, did not have independent guidance; the three warheads were intended to reenter the atmosphere in a shotgun pattern with the target at the center. Since such warheads cannot be aimed at separate targets, they do not alter the exchange ratio and do not provide any first-strike advantage. Their advent was therefore not in itself regarded as destabilizing.

In the early 1960's there was considerable debate over the appropriate size of the Polaris fleet. The Navy originally sought 48 ships, and the final decision was to build 41. One factor limiting the number of submarines is the problem of manpower recruitment. Nuclear-submarine duty, which involves 60-day underwater cruises, calls for a certain type of person who is not easy to find and who must be highly trained. Normally each vessel has two crews of about 140 men who go on alternate patrols.

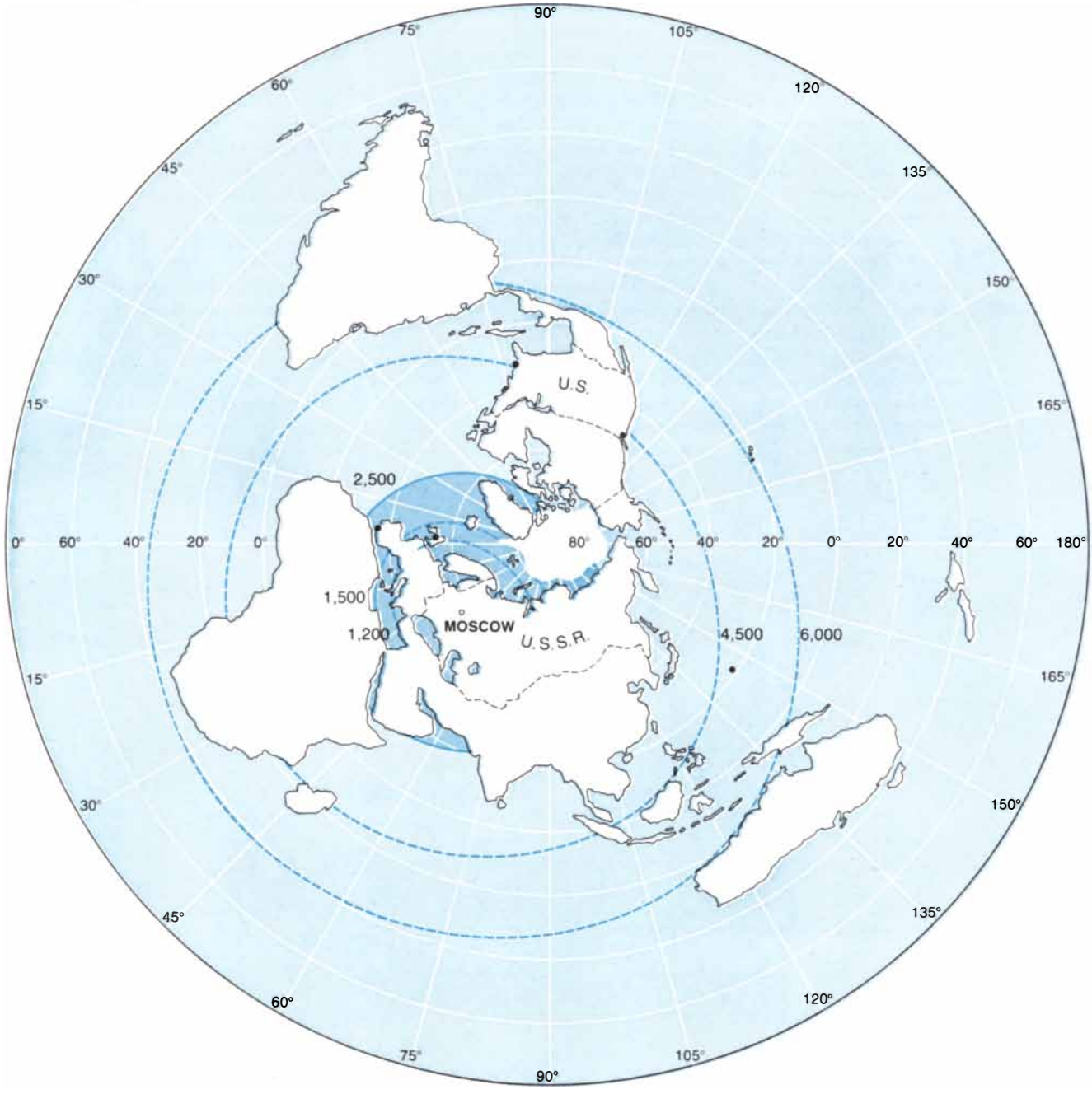
By the end of 1966 all 41 Polaris submarines were operational; eight carried *A-2* missiles and 33 were eventually



alternate crews consists of 14 officers and 126 enlisted men. The transverse section (*AA'*) shows the relative sizes of the Polaris *A-3* missile (*left*) and the Poseidon *C-3* missile (*right*). All the 31 missile submarines in this class will be converted to the Poseidon MIRV system by 1976. The Russian *Y*-class submarines, although not as advanced, are believed to be roughly similar in design.

fitted out with A-3's. Thus the force carried a total of 1,712 warheads, but since the triplet warheads cannot be aimed separately, 656 was the maximum number of separate targets that could be hit. Of course, not all of these submarines can be kept at operational stations at all times. In general a submarine spends 60 days at sea and 30 days in port for maintenance. In addition the submarine might take five or more days to move from the U.S. to its launch point and the same period to return. If a submarine wished to avoid detection by moving quietly and therefore slowly, the travel time would be even greater. Thus the number of submarines at launch stations at any one time could be reduced to some 20 to 25 ships. The situation is improved by using forward bases (on Guam, at Holy Loch in Scotland and at Rota in Spain), which reduces the time needed to reach launch stations from five or six days to one or two days.

With a range of 2,500 nautical miles, a submarine-launched missile can hit Moscow from most of the North Atlantic (inside an arc extending from the tip of Greenland to North Africa), from the Mediterranean and even from some parts of the Indian Ocean, a total sea area of about six million square miles [*see illustration below*]. The sea area from which a submarine-launched missile could hit important targets other than Moscow, say targets only 200 miles inside the U.S.S.R., is even larger [*see illustration on opposite page*]. One high-ranking Navy officer reported in 1964 that a Polaris submarine equipped with the A-3 missile could operate in 15 million square miles of ocean area while

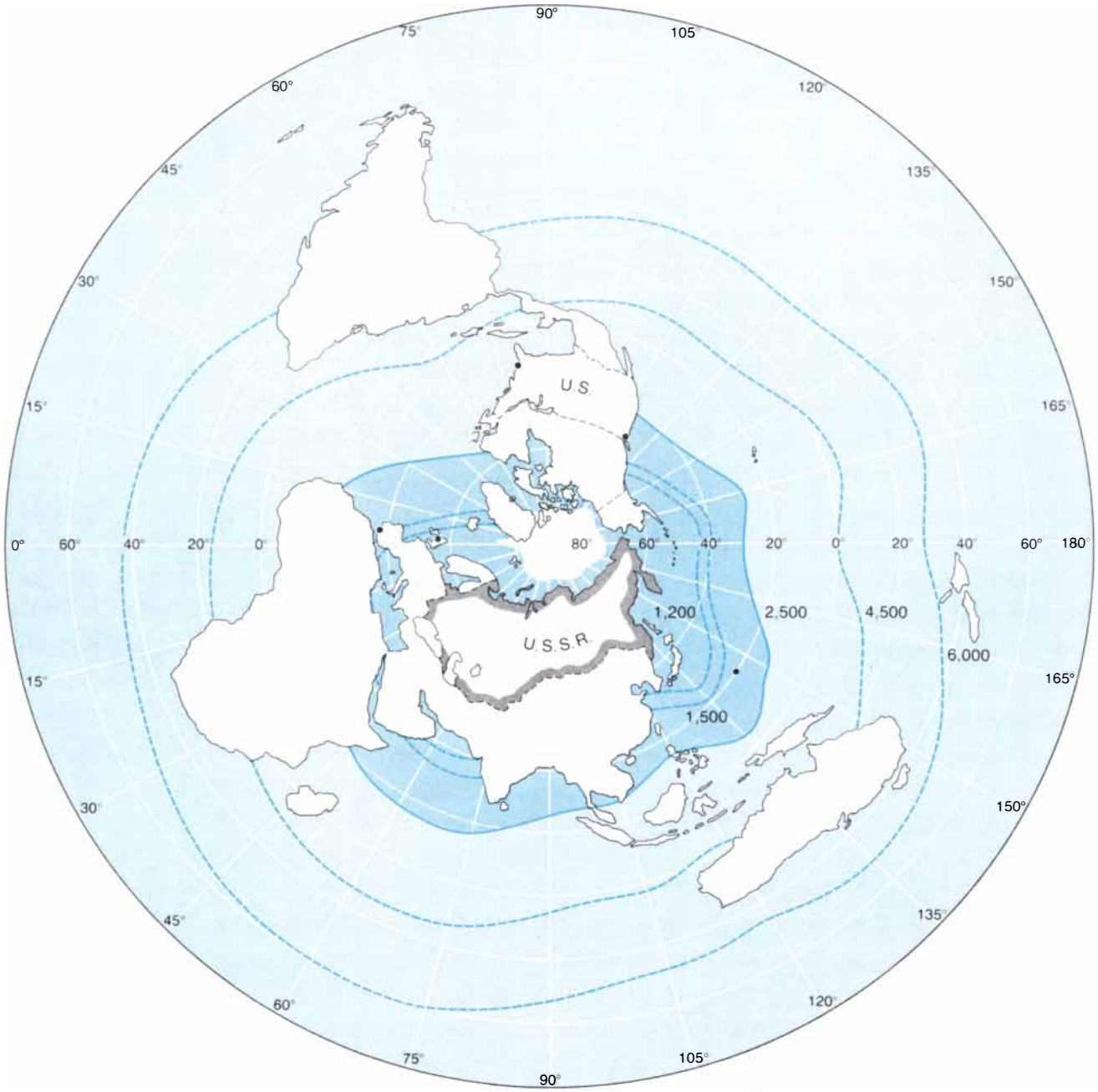


RANGE OF U.S. SUBMARINE MISSILES defines the operational sea areas from which submarine-launched ballistic missiles (SLBM's) could hit strategically important targets in the U.S.S.R. In the map at left the target is assumed to be Moscow; in the map at right the targets are assumed to include population centers and industrial complexes within 200 miles of the border of the U.S.S.R. (*gray areas*). The contours in the map at left are actually circles; their somewhat misshapen appearance is caused by the polar projection used in plotting the map. The contours in the map at right are concentric with the border of the U.S.S.R. The solid colored

covering its targets in the U.S.S.R. No land target anywhere in the world is inaccessible from attack by the *A-3* missile.

Although mobility provides a submarine with the tremendous advantage of improving its survivability, it creates a new problem: the determination of its location at the moment when the missile is to be launched toward its target several thousand miles away. Unless the missile is provided with some means of determining its position during flight or with a terminal-homing capability, the accuracy at the impact point can never be better than the uncertainty in the launch point. To determine the launch position calls for accurate submarine navigation, which is made more difficult by the requirement that in order to avoid disclosing its presence the submarine should not surface to determine its location. The attitude of the ship with respect to the vertical and true north at the time of the launch is also needed. When the missile force is being used for deterrent purposes, an accuracy greater than a few thousand feet is not needed; it is only necessary to be able to hit a large urban complex. Today this order of accuracy in locating the position and attitude of a submarine can be readily achieved. The U.S. has made tremendous advances in the development of inertial-navigation systems in recent years, and reasonably accurate position fixes can be obtained even after the submarine has been submerged at sea for many days.

The inertial-navigation system in a Polaris submarine is a complex system of gyroscopes, accelerometers and com-



contour in each map shows the range of the Polaris *A-3* and the nominal range of the Poseidon *C-3* missiles (2,500 nautical miles). The two inner broken colored contours in each case show the ranges of the older Polaris *A-1* and *A-2* missiles (1,200 and 1,500 nautical miles respectively). The two outer broken colored contours show the estimated ranges of the Navy's proposed undersea long-range missile system, which envisions two new generations of missiles, the ULMS 1, with a range of 4,500 nautical miles, and the ULMS 2, with a range of 6,000 nautical miles. Black dots denote home ports and forward bases of the U.S. missile-submarine fleet.

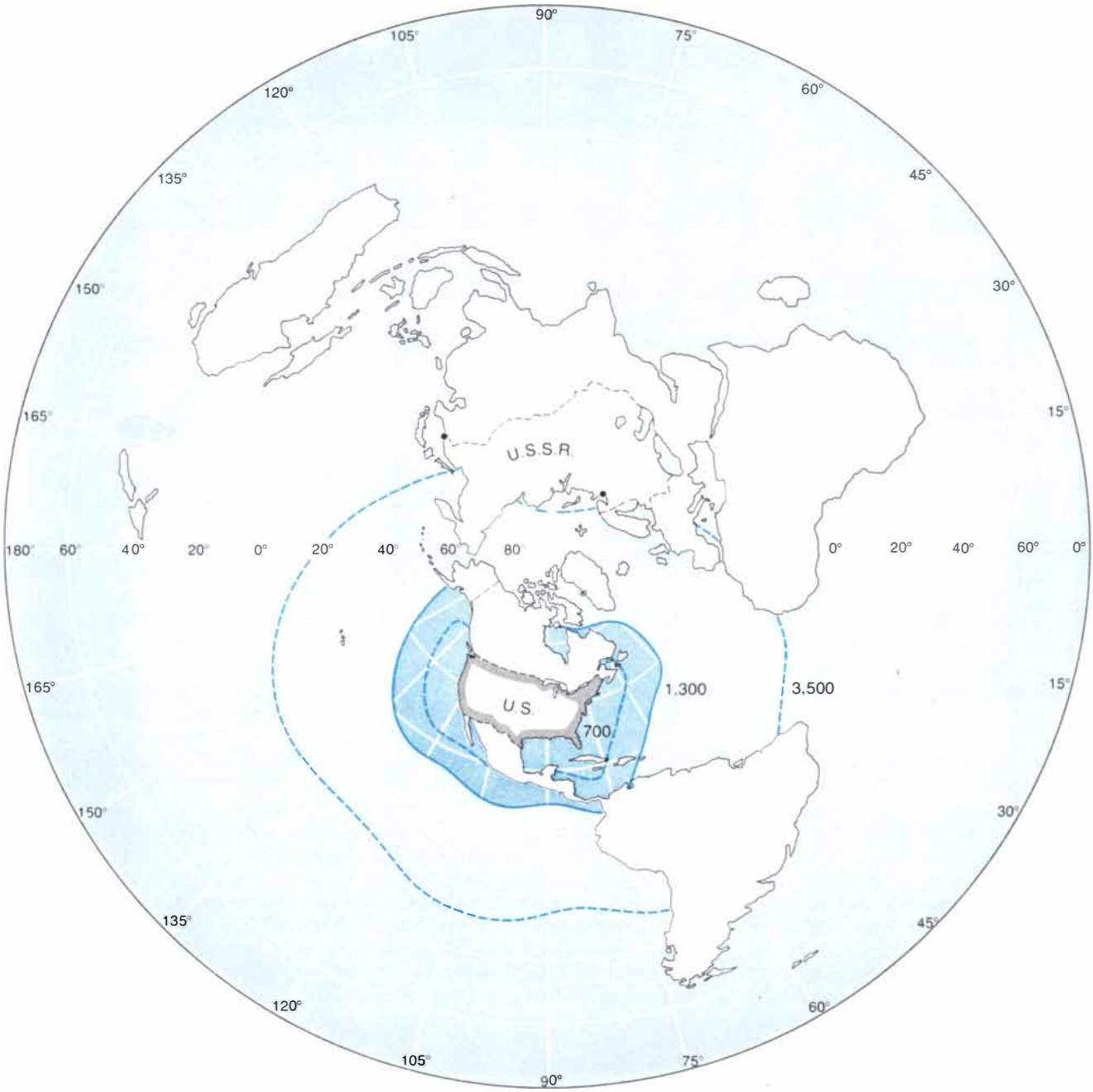
puters that relate the movement and the speed of the ship in all directions with respect to true north. If an initial position is known, then this system will provide continuous data on the ship's position. For an absolutely stationary submarine, or one whose motion can be corrected for, inertial sensors can determine without external data the vertical, the true north, the latitude and all velocity components by inertially sensing the earth's gravitational and rotational vectors, but there is no way of determining the longitude by inertial means. Submarines that have been voyaging at sea for protracted periods and whose inertial-navigation errors may have become unacceptably large can, by trailing an antenna while they are still submerged, get a radio "position fix" from navigation satellites or land-based transmitters. It may also be possible to locate a submarine by reference to accurately known geographical landmarks on the ocean bottom such as seamounts. In sum, the present technology has advanced to the point where the location and attitude of the submarine could in principle no longer be the critical factor in obtaining missile accuracies down to less than an eighth of a mile.

A deterrent force must also be able to receive communications from national command centers. Direct command and control originating with the President and with many verification checks is vital to prevent unauthorized launching; it is also essential that command authorities be able to communicate in times of crisis with the submarine captains without fear of interference by the other side. Otherwise communication would

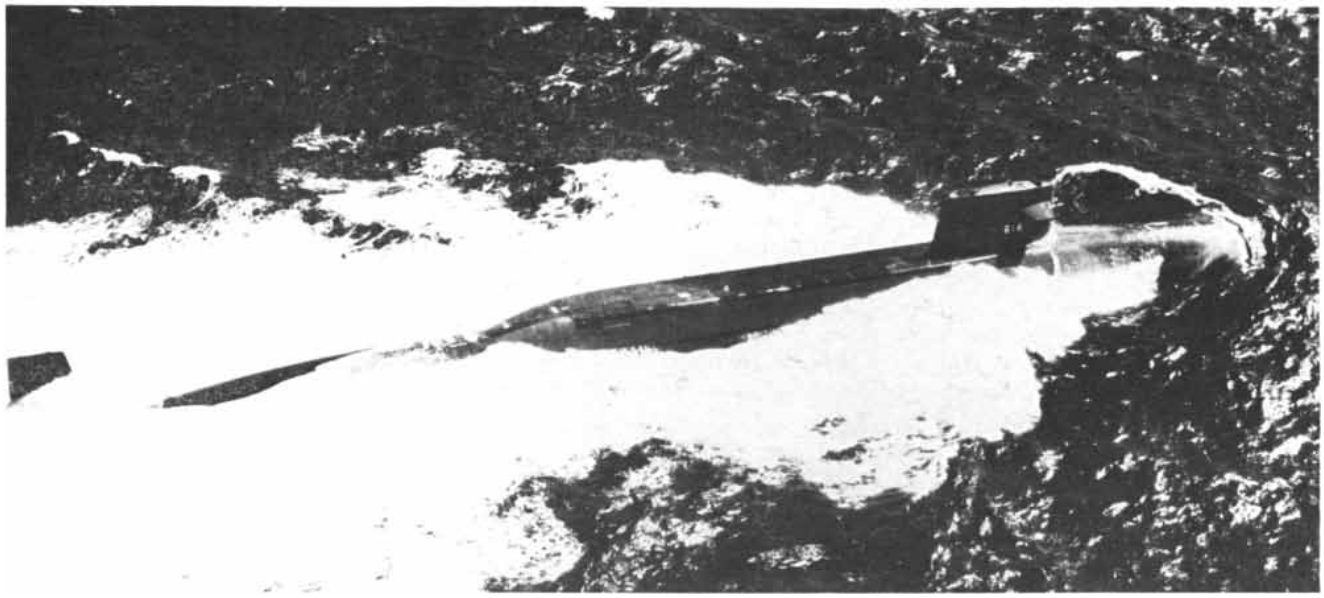
RANGE OF RUSSIAN SUBMARINE MISSILES is estimated in these two maps in order to show the operational sea areas from which their SLBM's could hit targets anywhere in the U.S. (*left*) and targets within 200 miles of the U.S. border (*right*). Alaska and Hawaii are excluded from consideration. Again the solid colored contour in each map shows the present range of the Russian *Y*-class missiles (1,300 nautical miles). The inner broken colored contour in each case shows the range of the previous generation of Russian SLBM's (700 nautical miles). The outer broken colored contour in each case shows the range of a new Russian missile that

be the Achilles' heel of the submarine deterrent. There are a number of means of communicating with a submarine, at least one way from shore to ship, that do not require the submarine to surface. Very-low-frequency (VLF) radio waves can penetrate a short distance into the water, so that a receiving antenna does not need to be exposed at the surface. Moreover, the submarine can operate at a considerable depth, since it can trail an antenna as much as several hundred feet above its deck. The U.S. has a number of land-based VLF transmitters at various locations around the world for communication with Polaris submarines, and more recently an airborne VLF system has been devised in order to eliminate the possibility that the fixed land-based stations would be destroyed in a surprise attack.

The use of satellites for relaying messages to submarines provides an alternative means of communication. Recently much research has been devoted to extremely-low-frequency (ELF) waves, which can penetrate even deeper into the water. The Navy project named Sanguine proposed to set up a vast antenna for this purpose in Wisconsin. The data rate of such a system would be quite low, but it would be adequate for command-communication purposes. The project has run into difficulties with local residents because of the large antenna currents and the potential hazard to living things. For communication from a submarine to the command center the problem is more difficult; it calls at the very least for trailing an antenna close to the surface and must in any case be avoided in order to prevent disclosure of



has recently been tested; this missile, however, is believed by U.S. strategic planners to be too long to fit readily in the present *Y*-class missile submarines. The black dots indicate the shipyards of the Russian missile-submarine fleet; the U.S.S.R. has no forward bases for its missile submarines. In general the greater distance from the Russian missile-submarine fleet's home ports to their operational launching areas, combined with their lack of forward bases, means that their submarines waste more time in getting "on station." Hence the U.S.S.R. requires more missile submarines than the U.S. to maintain the same deterrent force operational at any one time.



U.S. MISSILE SUBMARINE, the U.S.S. *Lafayette*, namesake of the advanced 616 class of Polaris vessels, was photographed while she cruised in the Atlantic Ocean during her builder's sea trials in April, 1963. The *Lafayette* was built in the shipyards of the Electric Boat Division of the General Dynamics Corporation in Groton, Conn. As originally deployed, this submarine was armed with 16 solid-fueled Polaris *A-2* missiles; it is scheduled to be converted to Poseidon *C-3* missiles with MIRV warheads sometime before 1976.

the submarine's location to listening enemy radios. Fortunately such communication is not essential to the viability of the submarine deterrent force.

By the end of 1966 the U.S. submarine-missile force together with its support systems was by itself more than adequate to deter any nuclear attack on the U.S. It had more than enough missiles and warheads to devastate the U.S.S.R. even when only a fraction of the submarines were on station. It could operate in ocean areas on all sides of the U.S.S.R., and the Russian ASW capability was quite rudimentary, with virtually no ability to "draw down" the size of the U.S. fleet. At that time the Russians had no ABM system deployed.

Military technology did not, however, stand still. The need to operate in restricted sea areas close to the northern coast of Europe and the Mediterranean in order to reach Moscow and other interior Russian cities created fears that someday ASW measures might become a threat. More important, concern that the U.S.S.R. might deploy a large ABM system capable of coping with our missile-submarine force was becoming more acute every day. The Russians were in the process of deploying an ABM defense around Moscow, using a large interceptor missile estimated to have a single-warhead yield large enough to destroy all three warheads of the Polaris *A-3*. In addition they were deploying radars and defensive missiles in the "Tallinn system" widely throughout the U.S.S.R. Some "worst case" analyses of U.S. planners, particularly during the early phases when factual information was limited, postulated that these facilities were for ABM defense. Later, as more data became available, the predictions were scaled down to the effect that the Tallinn system was an antiaircraft defense that could perhaps be upgraded to provide an ABM capability. (Now even this upgrading is not considered practicable by most experts.)

As a result research and development proceeded on a next-generation missile for the Polaris submarine that would give increased future assurance of penetrating any ABM defense and would at the same time give the submarines enough flexibility to operate at greater distances from the U.S.S.R. To increase either the payload or the range significantly called for a larger missile, and the resulting Poseidon missile required enlarging the launching tubes in the Polaris submarines, a costly and time-consuming task requiring 13 or more months. Since many of the submarines were due for overhaul in any case, however, the two shipyard activities could be combined with a minimum loss in operational readiness for the fleet as a whole. The cost of converting a Polaris submarine to carry the advanced Poseidon missiles is on the average $29 million, with another $38 million for normal submarine overhaul and replacement of the nuclear fuel.

The new Poseidon missile is about twice as heavy as the Polaris *A-3* and has a payload about four times as great. Although its nominal range of about 2,500 nautical miles is the same as that of the *A-3*, a trade-off between range and payload is always possible, so that the potential range of the Poseidon is somewhat greater. The new missile incorporates MIRV technology, that is, the ability to disperse many warheads aimed at separate targets. The technique developed for this purpose employs the "bus" approach, in which shortly after burnout of the propulsion stages the missile's final stage (the bus) is aimed at a first target point and releases a warhead, which then follows a ballistic trajectory to that target while the bus is redirected toward a second aim point. The same procedure can be repeated until all the warheads have been sent to individual targets. If a single target is to be attacked, then MIRV technology allows the warheads to approach the target at widely spaced intervals and on different trajectories so that no more than one warhead can be destroyed by a single ABM interceptor. The Poseidon is reported to be capable of carrying 14 warheads, each with a yield of about 50 kilotons, several times the yield of the bomb that destroyed Hiroshima. Warheads can be traded off for either ABM penetration aids or increased range. The nominal complement is usually taken as being 10 MIRV warheads. Department of Defense officials have repeatedly stated that these warheads do not have the accuracy-yield combination to provide a first-strike capability against hardened

silos (a "circular-error probability" of about an eighth of a mile or better would be needed with that yield), but the Russians might still be concerned on this score.

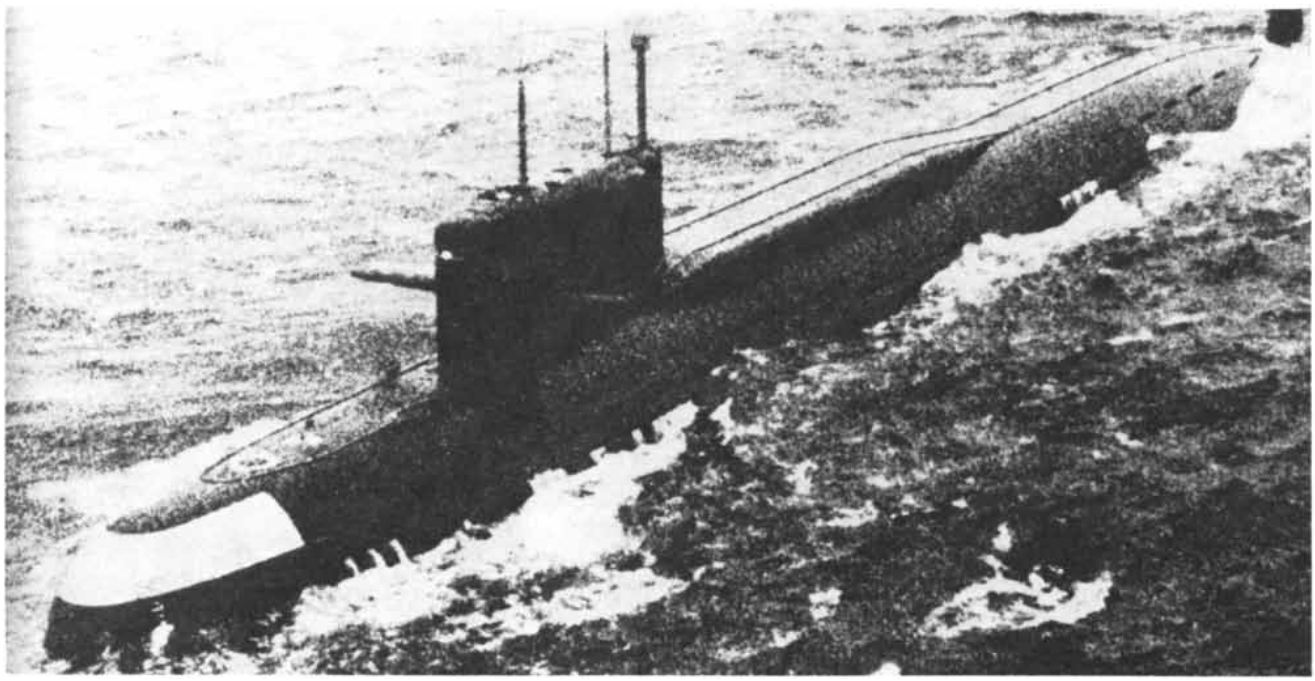
As with the Polaris *A-3*, a Poseidon missile with a range of 2,500 nautical miles can launch warheads at Moscow not only from large areas of the North Atlantic and the Mediterranean but also from some parts of the Indian Ocean, a total area of about six million square miles. Targets within 200 miles of the border of the U.S.S.R. can be reached from some 15 million square miles. These large ocean areas present great problems for any possible future ASW system. To deploy detection and tracking systems throughout these waters is a prodigious, if not impossible, task. Furthermore, on short notice these areas might be somewhat enlarged if it ever became critical by reducing the Poseidon payload, either by eliminating penetration aids or by cutting down the number of warheads in each missile.

The first Poseidon missile was tested in August, 1968, and the development of the entire Poseidon system was completed two years later. The first Polaris submarine went to sea with Poseidon missiles in March of last year. At present about 10 submarines have been converted. The program calls for modifying 31 submarines to carry these new missiles, leaving 10 to be equipped with the older *A-3*'s. When the program is completed in 1976, the U.S. submarine force will be able to launch 5,440 warheads at 5,120 separate targets. It should be possible to keep considerably more than half of these submarines on station at all times, and in times of crisis the operational readiness can be stepped up if necessary. This is an awesome force, capable of overwhelming even a massive ABM defense system. There is, of course, no evidence that the Russians have any intention of building a large ABM system with nationwide coverage, and it is highly likely that such a system will be precluded in the first stage of a strategic-arms-limitation treaty.

Even one missile submarine can launch 160 warheads at separate industrial centers in the U.S.S.R., an attack that the Russians could not afford even if the U.S. had been annihilated. This means that any ASW system would have to be able to eliminate almost instantaneously every single submarine, a herculean task. Today it is difficult, if not impossible, to destroy even a single submarine that follows skilled evasion tactics. Yet if ABM defenses were forbidden by treaty, an ASW system that has still to be devised would be the only threat to the submarine deterrent. That is one reason why an ABM agreement at SALT would by itself be such an important gain to the security of both the U.S. and the U.S.S.R.

The U.S.S.R. has always lagged considerably behind the U.S. in the development of nuclear submarines and SLBM's. The first Russian nuclear submarine was built about four years after the *Nautilus*, and Admiral Hyman G. Rickover, the director of the U.S. nuclear-submarine program from the beginning, has made it clear he believes the Russian submarines are technically inferior to the U.S. ships. The first Russian ballistic-missile submarines were diesel-powered and therefore had limited endurance and cruising range. Their first nuclear-powered missile submarine carried only three missiles with a 300-nautical-mile range, which in later models was extended to about 700 nautical miles.

By the late 1960's it must have been obvious to military planners in the U.S.S.R. that their land-based ICBM's would become increasingly vulnerable to the U.S. MIRV's, which were then under development and which had been publicly justified as providing an improved counterforce capability. The Russian deterrent needed shoring up with a more effective SLBM force, whose value had been demonstrated by the U.S. In 1966 the U.S.S.R. launched its first *Y*-class submarine, which carries 16 missiles with a reported range of 1,300 miles. This class of vessels was similar to the Polaris submarines, which the U.S. had put into operation seven years earlier. All the Russian SLBM's deployed so far have had storable liquid fuels,



RUSSIAN MISSILE SUBMARINE, a representative *Y*-class vessel, is remarkably similar in appearance and general dimensions to the earlier U.S. Polaris designs. The ship carries 16 liquid-fuel missiles with a reported range of about 1,300 nautical miles. Each missile is armed with a single nuclear warhead with a yield of about one megaton. This photograph was released by the U.S. Navy in 1970.

whereas all the U.S. missiles have had solid fuels. The Russians apparently decided to continue with the Y-class design and began building submarines at a rapid pace, initially at the rate of six to eight per year and currently at nine per year. Two shipyards are engaged in this work, one at Severodvinsk on the Arctic Sea and one in Siberia on the Pacific. At present the Russians have about 26 Y-class submarines operational and another 16 under construction.

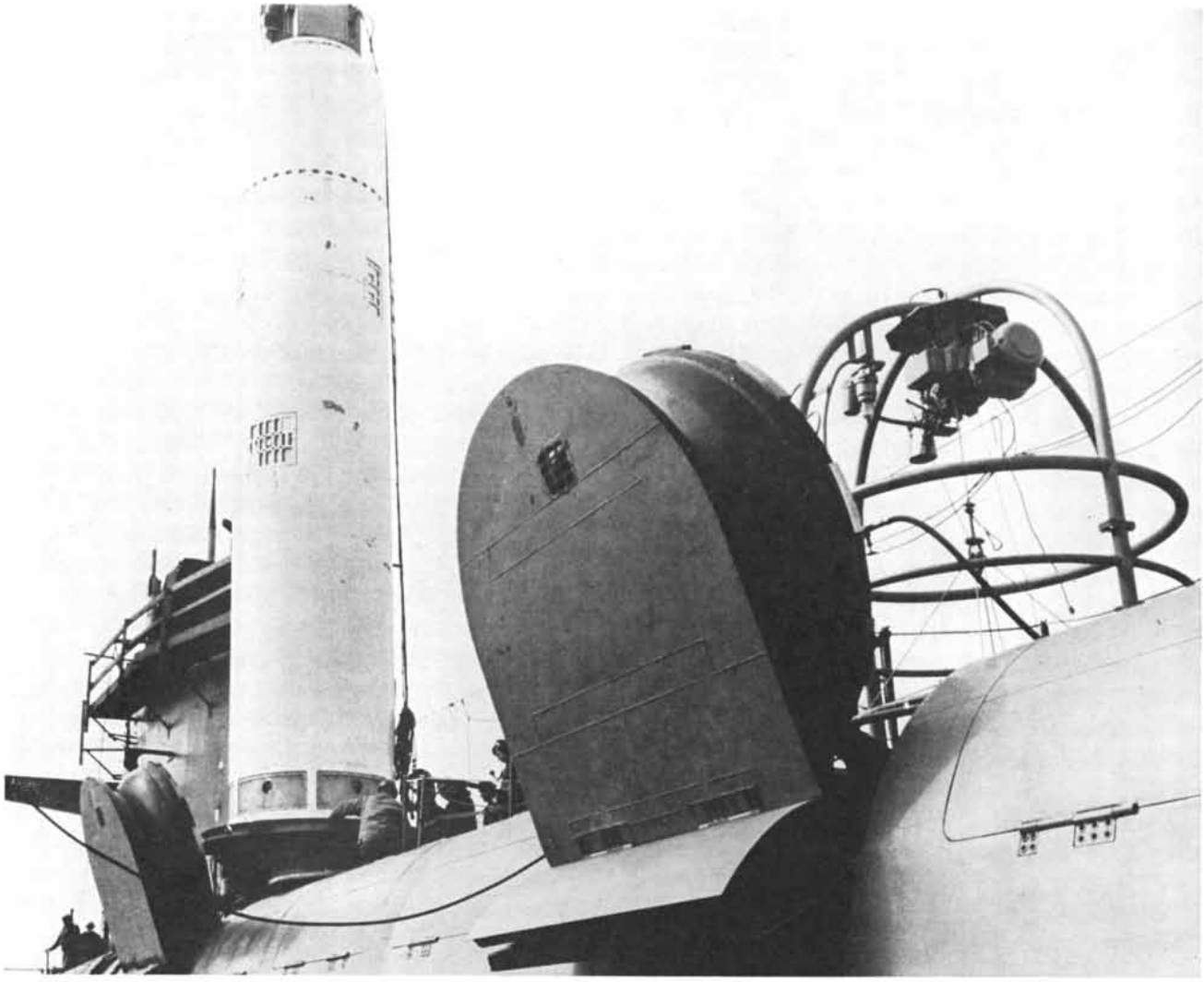
Although the Russians have tested a new missile with a range of about 3,500 miles, John Stuart Foster, Jr., chief research scientist for the Department of Defense, recently reported that this missile was so long that he did not believe the Y-class submarine could be modified to launch it. The Russians have never even tested multiple warheads on their submarine missiles, let alone MIRV's. Their missiles are each armed with a single warhead with a yield of about one megaton. These missiles have no capability for attacking our ICBM silos, but it has been postulated that they might be employed to attack our bombers on the ground and our command and control centers, using a depressed trajectory to achieve the necessary surprise. The missiles have not, however, been tested in this mode, and this approach would, in any case, entail a reduction in their already limited range.

The U.S.S.R. will have a slightly larger ballistic-missile submarine fleet than the U.S. when it completes those vessels now under construction (42 to 41). Even then, however, the capabilities of the Russian fleet will be far inferior to those of the U.S. Polaris fleet. President Nixon in his 1972 State of the World Message said that "our missiles have longer range and are being equipped with multiple independently targetable warheads. Moreover, our new submarines are now superior in quality." The shorter range of the Russian missiles requires that their submarines operate fairly close to the U.S. coast in order to be able to strike inland U.S. targets; this makes the Russian submarines potentially more vulnerable to a U.S. ASW system [*see illustration on page 20*]. On the other hand, the population centers and industrial complexes on the east and west coasts of the U.S. can be reached from much larger ocean areas, and these targets would be quite satisfactory if the SLBM's were to be used for deterrent purposes [*see illustration on page 21*]. The restricted range would be a serious factor only if the SLBM's were to be used against our bomber bases or missiles in the interior of the U.S.

There are other reasons why the parity between the U.S. and the U.S.S.R. in operational submarines cannot be evaluated on numerical grounds alone. Since bases in the U.S.S.R. are farther from the operational launching areas and since the Russians have no locations available for forward bases, more time is wasted getting submarines on station and it takes more submarines to maintain the same deterrent force operational at any one time. It would take Russian submarines a minimum of six days in the Atlantic and eight days in the Pacific to reach the nearest launch stations, so that the transit time to and from home ports, in many cases a quarter to a third of the duration of the patrol, seriously degrades the operational readiness of the Russian fleet. This disadvantage can be only partly alleviated by using submarine tenders for maintenance and crew exchange at sea. Moreover, in any East-West comparison the small British and French missile-submarine fleets, each of which may eventually consist of four submarines and 64 missiles, must be added to the U.S. total. Thus, whereas the Russians now have an adequate missile-submarine deterrent, their fleet is markedly inferior to that of the U.S. and its allies and provides no threat to the U.S. deterrent.

As the first phase of SALT is drawing to an end, then, it is becoming universally recognized that ballistic-missile

ARRANGEMENT OF MISSILE SILOS in a Lafayette-class Polaris submarine is revealed by the open silo doors in this view. The submarine, the U.S.S. *Sam Rayburn*, was photographed at Newport News, Va., at about the time of her commissioning in December, 1964.



CONVERSION of the Polaris submarine U.S.S. *James Madison* to accommodate the new, larger Poseidon missile was undertaken in February, 1969, and was completed in June, 1970. This photograph shows the *Madison*, the first of 31 such missile submarines to be converted, being test-fitted for the new Poseidon missile system at the Groton shipyards of the General Dynamics Electric Boat Division.

submarines are the essential foundation of a secure and stable strategic balance. Under these circumstances it is only natural to investigate ways to still further improve submarine missile systems. The U.S. has had a research and development program in this area for several years, so that the developments at the frontiers of technology could be incorporated in the successor to the Polaris-Poseidon system. The particular system proposed by the Navy for this role has been called ULMS, for undersea long-range missile system.

One obvious way to improve the present submarine missile would be to extend its range, making possible the launching of missiles from larger ocean areas in all directions around the U.S.S.R. Increasing the missile's payload would allow the incorporation of more warheads per missile or additional ABM-penetration aids. Payload can, of course, always be traded off for range. A longer-range missile would reduce the time required to move from U.S. ports to launching areas and thereby reduce the need for overseas basing in order to maintain the submarines on station for a larger fraction of their cruising time. With a range of 4,500 nautical miles a missile could reach Moscow shortly after the submarine leaves the U.S., whereas with a range of 2,500 nautical miles at least three days' travel would be required. Thus for a 60-day cruise lengthening the missile range by 2,000 miles could increase the period of operational effectiveness by about 10 percent if forward bases were abandoned.

More advanced guidance systems employing terminal control (the ability to change the path of the warhead during reentry) are being developed to avoid interception by ABM systems or to improve accuracy. Higher accuracy obtained by this means or others is not required if missiles are to be used as deterrent weapons. Indeed, it might be construed by the other side as an attempt to attain a counterforce capability for a first strike. Although it is more difficult to acquire such a capability with a submarine missile system than with an ICBM because of the inherent limitations on the payload and the nuclear explosive yield and because of navigational complications, there are no scientific barriers to its achievement. The greatest technical restriction on the use of SLBM's for a counterforce first strike may lie in command and control. It may be feasible to preprogram and command the initial launchings, but there will inevitably be failures. The difficulties of directing subsequent firings to destroy the silos missed the first time appear to be virtually insurmountable.

The submarine itself can be improved

by making it quieter as it moves through the water, thereby rendering detection and tracking by passive acoustic techniques more difficult. Although increasing the speed of the submarine will make it somewhat harder for enemy ASW ships to follow it, the higher speed will also raise the level of noise produced by the submarine. In any case it is probably a losing proposition for a missile submarine to try to outrun an ASW vessel, which can always be designed to move faster. High speed will enable the submarine to reach its launch area more rapidly and thus reduce the time it spends in a nonoperational condition, but again the potential gains are not large, and they may be outweighed by the disadvantages. If all other factors are equal, it will require a bigger power plant and larger submarine, both of which will increase cost and detectability. Increasing the depth at which a submarine can operate is not particularly significant, at least for the depths that are likely to be achieved in the next generation of submarines; submarines can be detected acoustically and destroyed by nuclear depth charges or homing torpedoes at any reasonable depth.

FIRST SUCCESSFUL TEST LAUNCH of a Poseidon missile from a submerged submarine was carried out in October, 1970. The missile was launched from the *James Madison*. The Poseidon missile, like the earlier Polaris models, is expelled from its silo by compressed gas; its rocket engine is then ignited after the missile has cleared the surface of the water.

If space for more and larger missiles is needed to increase the destructive capacity and the ABM-penetrability of any single submarine, then larger submarines with bigger power plants will be required. The larger the submarines, however, the fewer the ships that will be available for the same investment. Therefore if funds are limited (and they always are), this means a smaller fleet, which is more vulnerable to being wiped out in a simultaneous surprise attack. Thus there are many trade-offs in system design, and the final decision on a successor to the Polaris-Poseidon system should be based only on the nature of the specific threat.

In 1971 $104.8 million was appropriated for the advanced development of ULMS. Although this expenditure still left options open, it was a major step on the path toward procuring a specific new submarine system. This year the Department of Defense is seeking $977 million for ULMS. If that amount is authorized, the U.S. will be irrevocably committed to a large and very expensive new shipbuilding program. Unclassified details of the proposed ULMS program are still scarce, but it appears that the submarine in the system would be quite large (more than twice the size of the Polaris ships) and that it would be capable of launching 20 to 24 large missiles equipped with MIRV's. It is proposed to have a higher maximum speed and to incorporate the latest available silencing techniques, although these two objectives are competitive.

The ULMS program has been divided into two parts. The first stage (ULMS 1) would involve a new missile with a range of about 4,500 nautical miles capable of being deployed in the present Polaris submarines as well as in any new vessel. The second stage (ULMS 2) would include the development of the new submarine and a still more advanced missile with a range of about 6,000 miles that would be too big to be

substituted for the Poseidon in the existing Polaris ships. A maneuvering reentry vehicle (MARV) is also being developed for the new ULMS missiles. According to one estimate, the total cost of a program for 30 such ULMS vessels would be $39.6 billion [*see illustration at right*].

So far no convincing case has been made for the need to proceed with a replacement for the Polaris-Poseidon system and for making a commitment to a new large, high-speed submarine. Russian construction of SLBM's is no justification for ULMS; the Russian missile submarines do not in any way threaten the Polaris deterrent. Numerical superiority in launchers is meaningless; all authorities agree that the U.S. is far ahead qualitatively and can deliver from submarines about 5,000 warheads to fewer than 700 for the U.S.S.R. Even if we foolishly choose to race the Russians in the number of SLBM's, ULMS is certainly not the way to do it; each ULMS system will probably cost five or more times per missile launched than the Russian *Y*-class system.

The Poseidon with 10 or more MIRV's on each missile has a far greater capability than is needed to overwhelm any Russian ABM system that can be foreseen at present. Admiral Thomas H. Moorer, chairman of the Joint Chiefs of Staff, testified in February that "the Moscow ABM system even with improved radars and more and better interceptors could still be saturated by a very small part of our total missile force. In any event, the programmed Minuteman III and Poseidon forces, with their large number of reentry vehicles, provide a hedge against a future large-scale Soviet ABM deployment." Since such a large-scale ABM deployment will almost certainly be precluded by a first-stage SALT agreement, there is nothing in the ABM area that would require replacement of the Poseidon; in fact, even the Poseidon MIRV's will not be needed if a SALT ABM treaty is realized.

Therefore it is necessary to examine antisubmarine warfare to determine if there is anything that would currently justify the major ULMS step. Without going into a detailed analysis of possible ASW measures and countermeasures, suffice it to say that no evidence has yet been presented that the Russian ASW program could present a threat to the Polaris deterrent in the next decade. [An article on the ASW situation by Richard L. Garwin will appear in next month's SCIENTIFIC AMERICAN.] Admiral Levering Smith, director of the Navy's Strategic Systems Project, testified in 1969 that even the new generation of Russian ASW submarines will not be able to follow our Polaris submarines, and that the U.S.S.R. has no specific new ASW methods that would make the Polaris fleet vulnerable to attack. That is still true today. The U.S. has spent tens of billions of dollars on ASW efforts over the past 20 years and still does not have any system that could even begin to approach the kind of capability that would be needed to eliminate 20 to 30 missile submarines almost simultaneously. The Russians are far behind the U.S. in this area, and they have the serious geographical disadvantages of remoteness from and unavailability of land areas contiguous to the oceans in which their ASW systems would have to operate. Since the nature of the potential ASW threat to the Polaris-Poseidon system cannot even be foreseen at this point, ULMS, if built now, may be designed to cope with the wrong threat. The most obvious improvement to Polaris-Poseidon would be to increase the range of the missile in order to enlarge the ocean areas from which missiles could be launched. The deployment of such a new long-range missile might cost nearly seven billion dollars, however, and in any case, as the maps on pages 18 and 19 show, the Poseidon system already has a tremendous operational flexibility and is not threatened in its present launch areas.

Thus there are strong arguments for keeping both the ULMS missile and the ULMS submarine options in the research and early development stage. This would allow the exploration of all approaches, including smaller, slower but quieter submarines, and would avoid the making of a premature commitment to a large, expensive submarine and missile program. We must not fall into the trap of buying new military hardware just because we have made technological advances; there is no quicker way to price ourselves out of the security market. The submarine missile force is the backbone of our deterrent; its present strength and invulnerability obviate the need for its replacement for at least a decade.

RESEARCH, DEVELOPMENT, TEST AND EVALUATION	MILLIONS OF DOLLARS
SUBMARINE	$800
MANEUVERING REENTRY VEHICLE (MARV)	$600
ULMS 1 MISSILE	$2,100
ULMS 2 MISSILE	$1,500
PROCUREMENT	
SUBMARINES (30 AT AVERAGE COST OF $400 MILLION EACH)	$12,000
ULMS 1 MISSILES (ENOUGH FOR 500 LAUNCHERS AT AVERAGE COST OF $8 MILLION EACH)	$4,000
ULMS 2 MISSILES (ENOUGH FOR 600 LAUNCHERS AT AVERAGE COST OF $11 MILLION EACH)	$6,600
SUPPORT	
ULMS REFIT FACILITIES (ONE PER 15 SHIPS AT COST OF $1,800 MILLION FOR FIRST, $1,200 MILLION FOR SECOND)	$3,000
OPERATION AND MAINTENANCE (FOR 30 SHIPS AT AVERAGE COST OF $30 MILLION PER YEAR FOR 10 YEARS)	$9,000
TOTAL	$39,600

ESTIMATE OF TOTAL COST of the Navy's proposed ULMS program was prepared by Members of Congress for Peace through Law. The first stage of the ULMS program (ULMS 1) would involve a new missile with a range of about 4,500 nautical miles capable of being deployed in the present Polaris submarines as well as in any new vessels. The second stage (ULMS 2) would include the development of a new, larger submarine and a still more advanced missile with a range of about 6,000 miles. A special maneuvering reentry vehicle (MARV) is being developed for these new, larger missiles. The entry under research, development, test and evaluation for the submarine includes the cost of all nonmissile subsystems and integration work. The procurement estimates assume that the new program would eventually involve 30 new submarines. The support estimates assume a need for two special refit facilities for the ULMS ships and include crew-related costs for 10 years. The cost of the ULMS 1 missile program by itself is estimated to come to about $6.7 billion.

Markers of Biological Individuality

The rejection of transplanted organs has focused attention on the body's ability to distinguish foreign cells from its own. The cells of each individual are uniquely marked with protein

by Ralph A. Reisfeld and Barry D. Kahan

It is often observed that every human being is unique. Our superficial differences, however, scarcely hint at the differences to be found at the level of the genes and the chemistry of the cell. It is now becoming clear that these differences play an important role in enabling living organisms to distinguish between "self" and "not self" and to reject cells that are not labeled with the appropriate recognition markers. The advances in surgery that make it possible to transplant organs from one person to another have stimulated an intensive study of the factors that cause the recipient to reject foreign tissue unless the donor is a close relative. Criteria for matching unrelated donors and hosts have proved extremely difficult to establish. The reason, we now see, is precisely the genetic and chemical uniqueness of each individual. Where organ transplants have been successful it is usually because ways have been found to suppress the rejection mechanism that normally operates. The techniques of suppression, however, still depend more on good luck than on fundamental understanding. Here we shall try to summarize what has recently been learned about the unique markers of individuality carried by each cell.

Individual differences are readily apparent in anatomy and physiology. "Normality" can only be defined by statistical methods involving many individuals, from which one can arrive at mean values with limits of variation. Any given individual, however, must possess many physical characteristics that are outside the normal range but that are not necessarily pathological. Functional differences are subtler; for example, individuals commonly exhibit large differences in the ratios of the electrolytes present in blood serum or in the structure and function of the various enzymes that mediate specific biochemical reactions. It is clear that such differences are innate: they are demonstrable over periods of years and are only slightly influenced by the environment.

Individual peculiarities are sometimes tragically illustrated by the idiosyncratic reactions of patients to particular drugs. Fifty years ago Sir Archibald Garrod, in his pioneering book *Inborn Errors of Metabolism*, noted that "these [pathologic defects of enzyme chemistry] are merely examples of variations in chemical behavior which are probably everywhere present in minor degrees, and that just as no two individuals of a species are absolutely identical in bodily structure, neither are their chemical processes carried out on the same lines."

Since these individual differences can be preserved from generation to generation only by the maintenance of a complex polymorphic gene pool, that is, a pool of genes with many alternative expressions, they must serve some important purpose. Unless they contributed to the survival of the species they would have been eliminated long ago by the pressures of natural selection. Moreover, it is clear that individuality is not confined to higher organisms. As Thomas Humphreys and Aron A. Moscona of the University of Chicago have shown in their studies of specific aggregation of sponge cells, even primitive species can distinguish their own cells from others. Genetic variability enhances the ability of members of a species to survive changes in their external and internal environments. Thus by studying the factors that determine and express individuality one can hope to acquire insight into basic biological phenomena relevant not only to the role of isolated characteristics but also to the functioning of the organism as an integrated collection of cooperating processes.

One of the most valuable techniques for studying individuality pits the distinctive factors of one organism against those of another. It is easily achieved by transplanting a bit of tissue from a donor to a host. The host is then required to distinguish its own unique biological markers from those of the interloper. One might regard this as a test of parasitism: What prevents one individual from becoming an integral part of another?

Humans are not good subjects for transplantation studies because the uniqueness of each individual precludes getting a "typical" response. One would have to conduct many experiments with many individuals and interpret the results statistically, which would of course tend to obscure the individual factors one was setting out to study. The alternative is to use inbred strains of laboratory animals such as mice or guinea pigs, so that one has a better chance of isolating the chemical and biological determinants of individuality. In the 1920's Sewall Wright and Clarence Cook Little showed independently that by repeated brother-sister matings one can develop animal lines possessing a common pool of genes. This does not mean that each newborn individual contains the same set of genes but rather that each set is derived from the same limited pool. Wright and Little demonstrated that tissue grafts between members of the same line survived longer than grafts between members of two different lines.

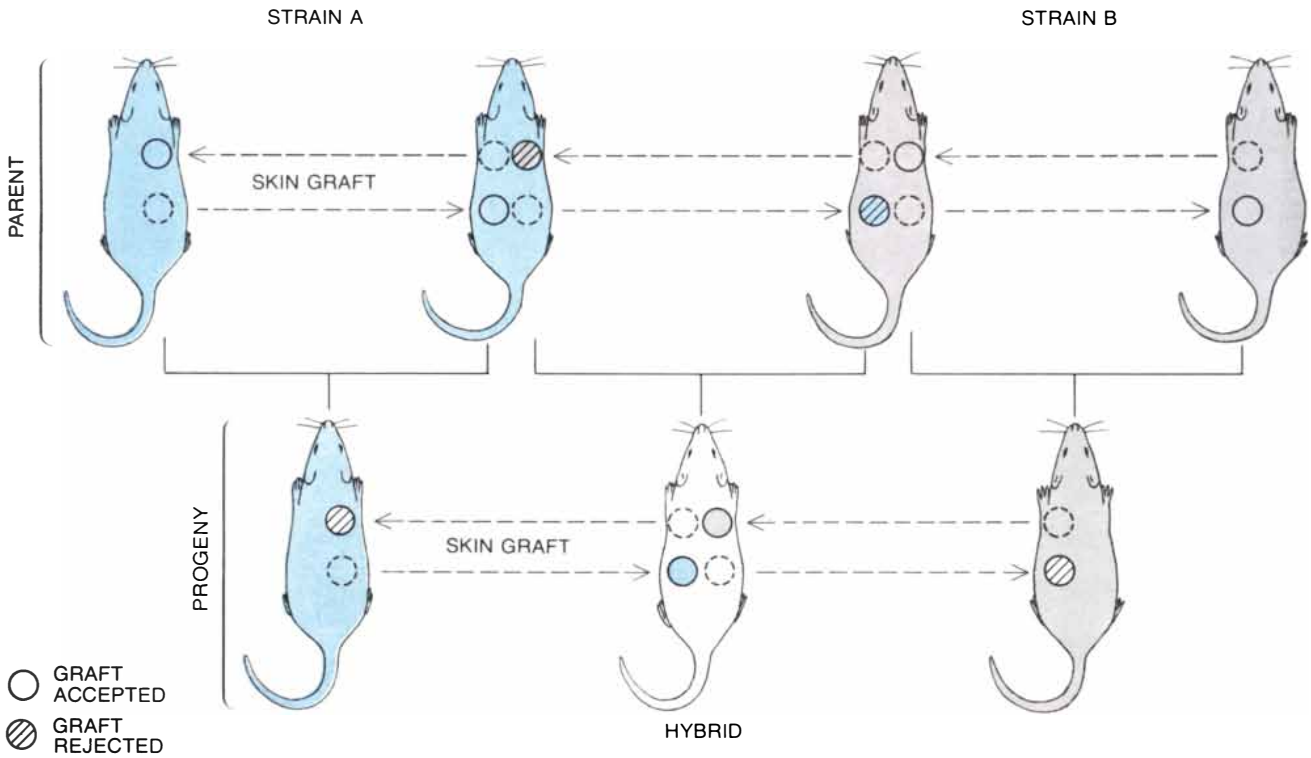
Little's colleague George D. Snell, working at the Jackson Memorial Laboratory, went further and showed that tumors obtained from one line of mice are rejected by hosts that differ from the donor strain by a factor subsequently designated *H*-2. The gene controlling the *H*-2 set of factors is transmitted by

simple Mendelian laws [*see illustration below*]. Grafts of normal tissue exchanged between members of the same inbred line (isografts) are accepted. Grafts exchanged between members of two different lines (allografts) are uniformly rejected within 14 days. Grafts transferred from parental lines to the first-generation hybrids of those lines are accepted, but grafts transferred from the hybrid offspring back to the parents are not.

Moreover, if the grafts are tumors rather than normal tissue, the response generated by the *H*-2 gene is strong enough to destroy the graft when the donor and the host possess different *H*-2 genes. Evidently the *H*-2 gene gives rise to strong markers on the surface of the cell that are readily recognized by cells of a different strain. In the absence of such a strong response tumors grow rapidly and outstrip the host's defenses. Snell also observed that when the grafts consist of normal tissue, they can be rejected by genetic factors weaker than those supplied by the *H*-2 gene.

Furthermore, working with congenic lines of mice (inbred lines differing from one another at a single genetic locus), Snell found that a number of weaker factors, distributed as loci in the genetic material different from the *H*-2 locus, could initiate the rejection of normal-tissue grafts, although rejection could take as long as 200 days. Differences at a second genetic site are not capable, however, of stopping the growth of tumor grafts when the donor and the host possess the same *H*-2 gene. The strong markers placed on cells by the *H*-2 gene in mice have their counterpart in man in markers traceable to the human leukocyte locus *A* (HL-*A*) gene. Thus transplantation experiments have revealed a number of discrete genes that control the production of markers, or distinctive factors, that determine the fate of foreign grafts.

Genetic markers of this type have long been exploited in the typing and matching of red blood cells for transfusions. The individuality markers that have to be accounted for when tissues are transplanted from one organism to another, however, are much more numerous and harder to classify because many genes are involved. When studies of tissue transplantation were extended to outbred populations of mice, it was shown that the number of alternative gene expressions associated with the strong factors that control rapid rejection of transplants in mice is probably no smaller than the number observed in man. A comparison of the data obtained in mice with observations in human populations led Jean Dausset of the University of Paris to suggest that the genes controlling the strong transplantation factors are collected in a single genetic region. This hypothesis was subsequently supported by the work of Ruggiero Ceppellini of the University of Turin and Jan van Rood of the University of Leiden. Ceppellini demonstrated that the rapid rejection of grafts exchanged between two human subjects depends on a single genetic locus; van Rood then found that the locus controls the production of factors that can be detected with HL-*A* antibodies. The region seems to be divided into two subregions, each of which determines a series of allelic, or alternative, factors. There are multiple alternatives in each subregion: the *D* and *K* regions in mice each have at least 10 alleles; in man the L-*A* region has at least 11 alleles and the region known as "Four" has at least 17. Thus a simple biological test, the acceptance or rejection of transplanted tissue, has uncovered a richly polymorphic genetic system in mice and men, providing a tool for attacking the problem of biological variability.



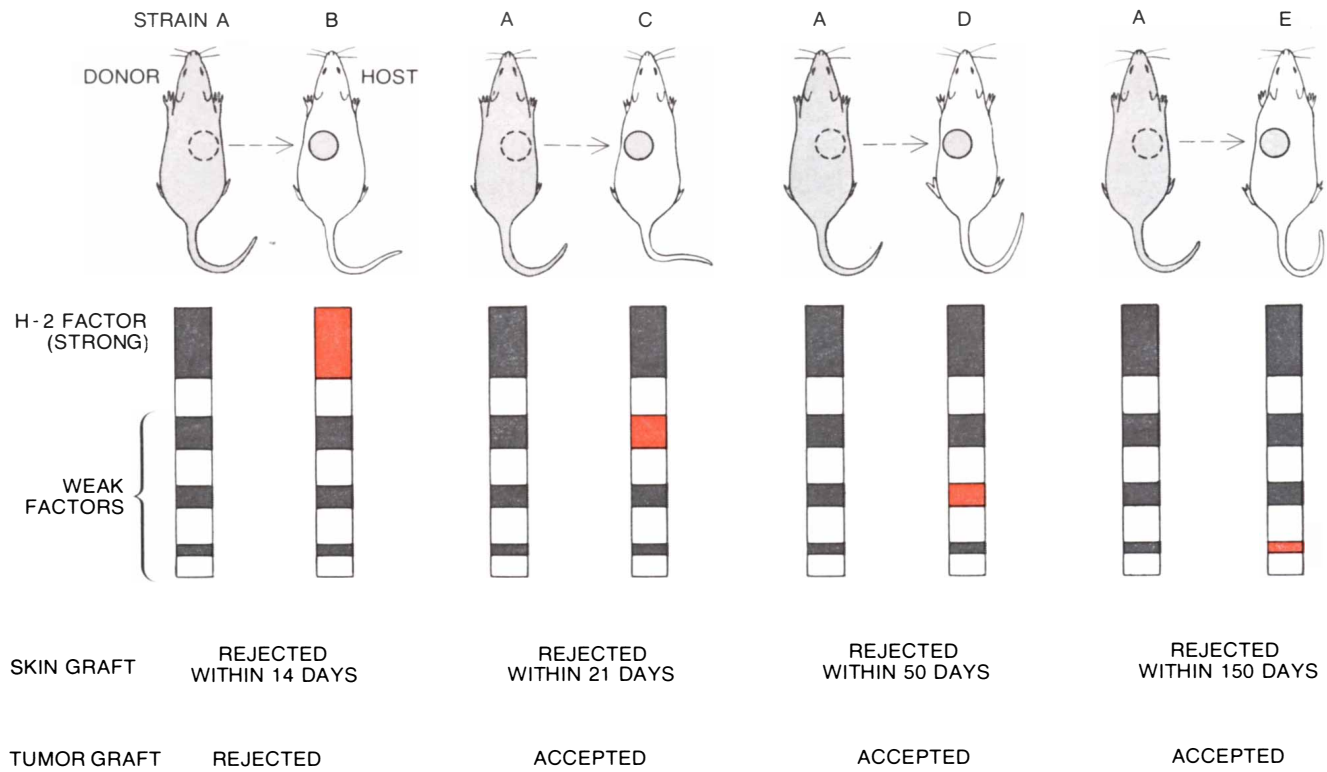
STRONG RESPONSE TO TUMOR TRANSPLANTATION in mice depends on factors controlled by a specific gene known as *H*-2. George D. Snell of the Jackson Memorial Laboratory demonstrated in the 1940's that the *H*-2 gene obeys simple Mendelian laws of inheritance. Tumor grafts exchanged between inbred lines of mice, such as between two mice of strain *A* or two of strain *B*, are accepted. Tumor transplants between mice of different strains, however, are vigorously rejected, being destroyed within two weeks. Hybrid offspring will accept grafts from either of the pure parental lines, but grafts in the reverse direction are destroyed.

The polymorphic marker system protects an individual from invaders that otherwise could not be distinguished from the self. Although data from human populations originally suggested that one out of every 10,000 unrelated individuals might possess the same strong transplantation genes, it has since been demonstrated that the frequency of identical genes is far rarer. In fact, Leo Loeb proposed many years ago that the possible combinations are infinite and that no exact matches can exist. The results of human transplantations offer little to refute Loeb's conjecture. Clinical success in organ transplants seems related less to genetic similarities between donors and recipients who are unrelated than to the impaired ability of the host to react to foreign markers.

What is the mechanism that enables the host to recognize and destroy foreign cells? There are many conceivable possibilities. For example, the host might provide a local environment in which foreign cells were deprived of essential nutrients. Plants are able to resist certain parasites by depriving them of substrates, for example polysaccharides, they need for growth. A parasite's success might depend on its ability to break up whatever such molecules may be provided by the host and to utilize their fragments. Another possibility, envisioned by Loeb, is that the growth of foreign tissue may result in the generation and release of specific substances that are toxic to the host. This might lead in turn to a nonspecific inflammatory response that would destroy the invader.

These and other hypotheses were finally ruled out some 20 years ago when P. B. Medawar and his colleagues at University College London demonstrated that the host's response to foreign tissue involves an immunological mechanism similar to the one that provides resistance to bacterial and viral infections. In his first group of experiments Medawar showed that the recipient of a graft develops a resistance that is specific for the donor: when the recipient is challenged with a second graft from the same donor, the transplanted tissue is destroyed even more rapidly than it was the first time. Medawar called this the second-set reaction [*see "a" in illustration on opposite page*]. Furthermore, the "memory" of the initial experience can be evoked by a second challenge graft no matter where it is placed on the recipient; in other words, the second-set response is system-wide, not local. Medawar then found that specific individuality markers are associated with every nucleated cell of an organism. Thus an animal can be "immunized" against foreign skin grafts by first injecting it with cells derived from other tissues of the same donor, for example cells from the spleen [*see "b" in illustration on opposite page*].

Finally, Medawar and two of his colleagues, Rupert E. Billingham and Leslie Brent, drawing on observations made by Ray D. Owen and a theory proposed by Sir Macfarlane Burnet, demonstrated that an animal of strain *A* can be made tolerant to tissue grafts from an animal of strain *B* by injecting the strain-*A* animal soon after birth with a suspension of spleen cells from the strain-*B* animal [*see "c" in illustration on opposite page*]. With this experiment Medawar and his colleagues proved that resistance or lack of resistance to foreign tissue is not a localized phenomenon but an immunological one. They later showed that the host's capacity to reject foreign tissues can be reconstituted by lymphoid grafts, which are known to restore immune reactions. This set of experiments left no doubt that the factors previously identified by genetic studies were those re-



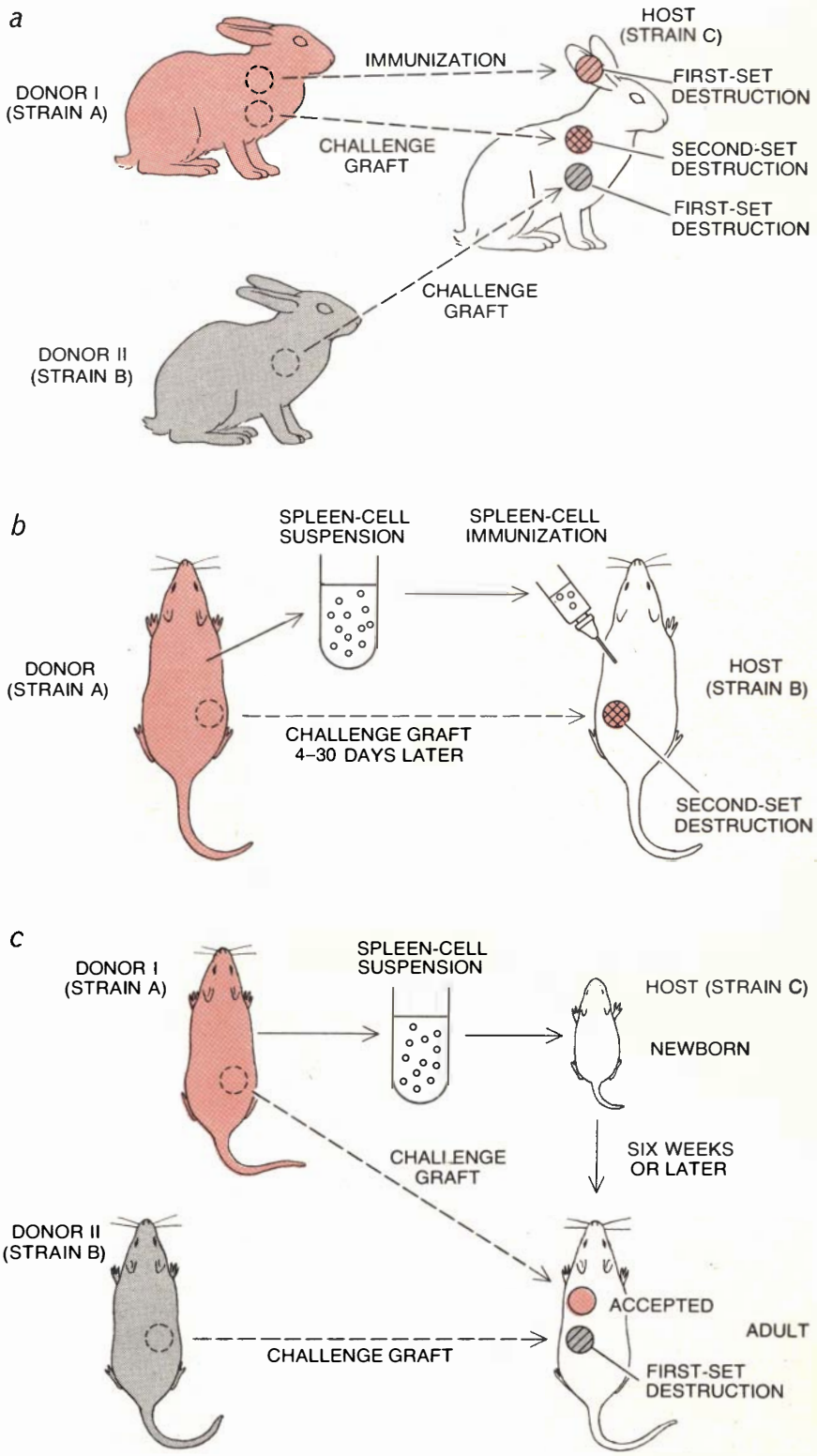
ADDITIONAL GENETIC FACTORS determine the survival rate of normal skin grafts but not the survival rate of tumor grafts. If an animal of strain *A* and an animal of strain *B* differ in *H*-2 (that is, strong) individuality factors, each will vigorously reject tumor grafts as well as ordinary skin grafts from the other. Animals that share the same *H*-2 factors will accept tumor grafts from one another. Because animals *C*, *D* and *E* differ from *A* in weaker transplantation factors they reject normal skin grafts at various rates.

sponsible for graft rejection, since alteration of the host's responses toward these factors made it unable to react against grafts.

To account for these findings and related ones Medawar proposed that all nucleated cells possess surface markers that act as "antigens" when the cells are transplanted to another organism. The antigens trigger an immune response in the host that makes it specifically resistant to the donor's tissues. Evidently these markers are functional components of the outer membrane of every nucleated cell, although the role they play remains to be elucidated. Presumably the host becomes alerted to their presence when they perform their natural functions, perhaps as cell receptors or as factors involved in growth and development.

James L. Gowans of the University of Oxford has shown that foreign grafts are met by "wandering" lymphocytes, white scavenger cells produced in the lymph glands; the lymphocytes "check" cell surfaces like watchmen in order to ensure that all constituents of the host organism bear its own markers and not those of an invader [*see illustration on next page*]. When the wandering lymphocytes discover foreign markers in a graft, they carry the alarm back to the lymph nodes and stimulate an immune reaction. The reaction consists in the proliferation of lymphocytes with receptor sites that can engage the markers of foreign cells as a key engages a lock and thus interfere with their functioning, ultimately causing the death of the grafted cells. There is also evidence that whole cells or fragments from the graft tissue are detached and find their way into the drainage channels of the lymphatic system, where they eventually reach the lymph nodes and directly trigger the proliferation of suitably designed lymphocytes. Presumably the transplant markers have a role in the metabolism of the cell that requires them to occupy an exposed position on the cell's surface, so that a reaction against them is lethal to a cell transplanted in a foreign host.

Working with crude subcellular fractions, a number of investigators obtained evidence some years ago that nearly all the strong antigens of the cell are located on the surface of the cell membrane. Goran L. Möller of the Royal Caroline Institute in Stockholm found that serums produced in a host by challenge with intact foreign cells contain antibodies that react with cell surfaces [*see illustration on page 35*]. That these



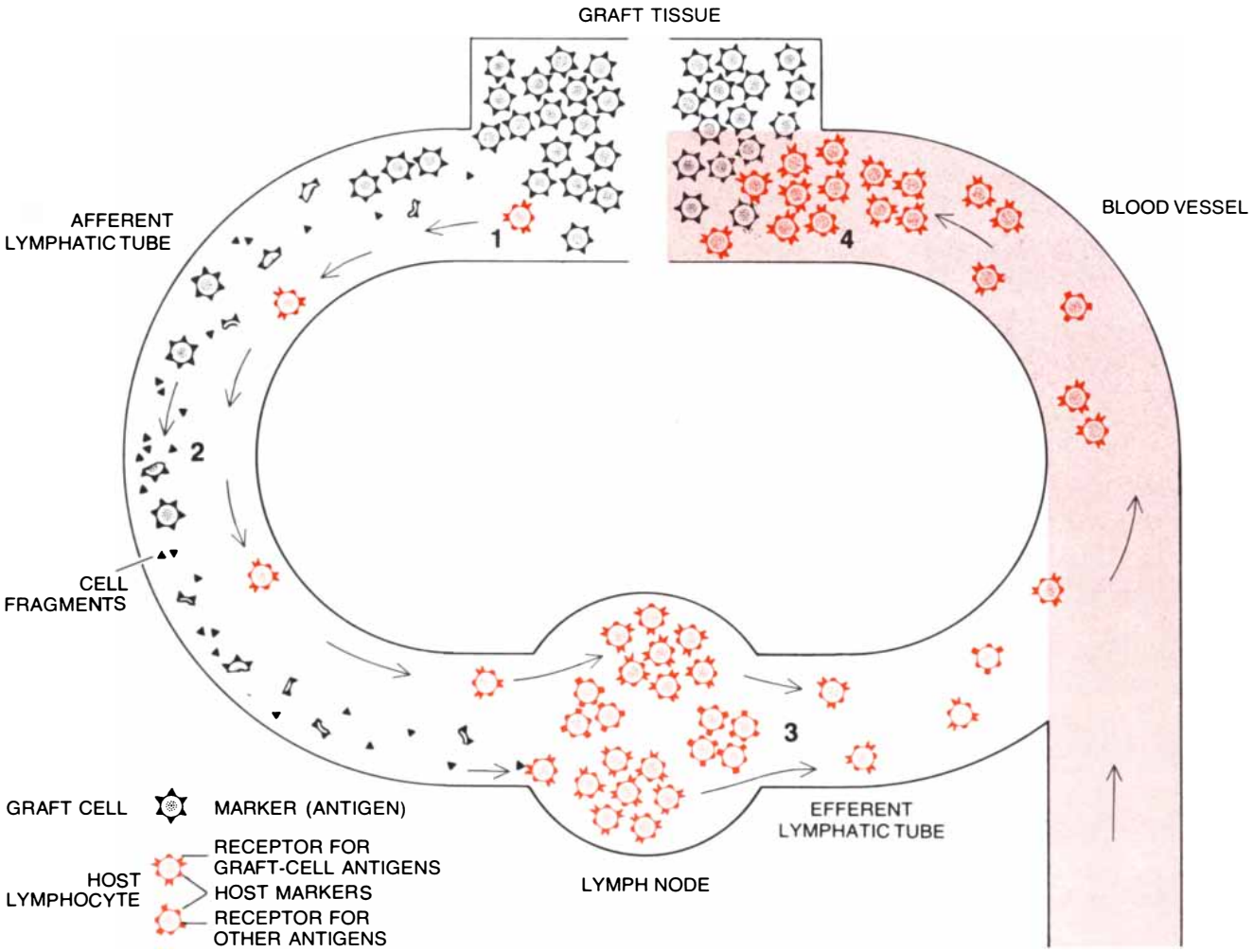
IMMUNOLOGIC NATURE OF TISSUE REJECTION was demonstrated by P. B. Medawar and his colleagues at University College London. His early studies showed (*a*) that a second graft from the same donor is destroyed more rapidly than the first. Medawar called this the second-set reaction; it shows that the host develops an "immunity" to a specific foreign donor. He then showed that immunization is not specific for the tissue involved (*b*). Animals exhibit a second-set reaction to a skin graft if they have previously been injected with cells derived from the donor's other tissues, such as the spleen. Finally Medawar and his group demonstrated that animals can develop immunologic tolerance (*c*). A mouse of strain *C* will accept a skin graft from a mouse of strain *A* if the strain-*C* mouse has been injected soon after birth with spleen cells from the mouse of strain *A*. The strain-*C* mouse, however, still retains its ability to reject challenge grafts from an animal of another strain, say strain *B*.

surface sites are indeed transplantation markers was proved in our laboratory at the Scripps Clinic and Research Foundation in La Jolla, Calif., by Soldano Ferrone, who recently demonstrated that solubilized and purified human individuality markers stimulate the production of antibodies that react with surfaces of intact cells.

Workers in a number of laboratories have been exploiting various techniques to isolate the individuality markers in a water-soluble form from the cell-membrane matrix with which they are associated. The task is complicated because the cell membrane consists of a lattice of lipoprotein molecules in which are embedded a variety of carbohydrates and proteins in addition to the individuality markers [*see top illustration on opposite page*]. Some of the proteins serve as structural components; others regulate the passage of nutrients, respond to environmental stimuli and effect the cell's associations with neighboring cells. Whereas the architectural components of the cell membrane tend to be fixed, some of the functional proteins probably float like icebergs in a "sea" of membrane lipids, so that they are at least partially exposed to the external environment. Michael A. Edidin of Johns Hopkins University has inferred that transplantation markers are among those in a dynamic state from the way they spread rapidly between membranes when two unlike cells fuse.

One current model of the cell surface visualizes membrane components held together either by strong covalent chemical bonds or by weaker noncovalent interactions, such as those provided by hydrogen bonds and salt linkages. Since the individuality markers appear to be mobile, it seems likely that they are associated with the architectural proteins and with other functional proteins through noncovalent interactions. One can only speculate on the role of these interactions. For example, by altering the expression of the individuality markers the interactions may protect the host against environmental agents that might alter the cell's identity. Fortunately it turns out that the antigenic activity associated with the individuality markers is independent of secondary interactions with other membrane components and depends solely on the marker's primary structure. Therefore it was possible for us to remove the markers from the membrane in water-soluble form and study their biological activity with a minimum of interference from other membrane components.

There are two general methods for re-



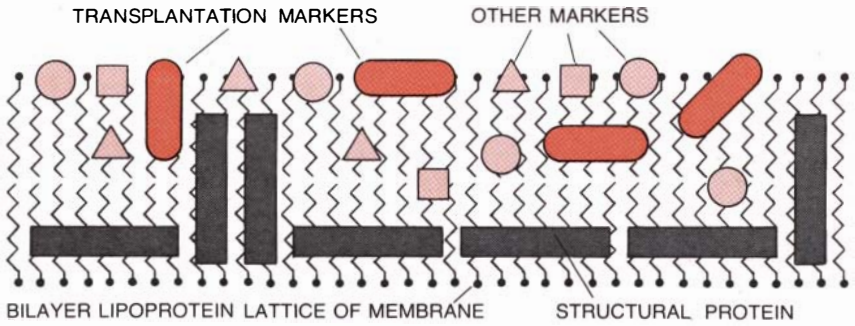
THEORY OF IMMUNOLOGIC RESPONSE to tissue transplants, proposed by Medawar, visualizes that cells carry individuality markers that another organism perceives as "antigens." Here the markers, or antigens, carried by the cells of a foreign graft are represented by black spikes. These are detected by the host's "wandering" lymphocytes, some of which are equipped with receptors that interact with the foreign markers in a lock-and-key fashion. These wandering lymphocytes (*1*) carry an alarm signal to the host's lymph nodes, stimulating the proliferation of more lymphocytes (*3*) equipped to destroy the invading cells (*4*). Cells detached from the graft, together with fragments carrying antigens, also enter the lymph drainage system (*2*) and carry the alarm message directly to the lymph nodes. Evidently lymphocytes are equipped with a variety of "keys" to engage any number of antigen "locks."

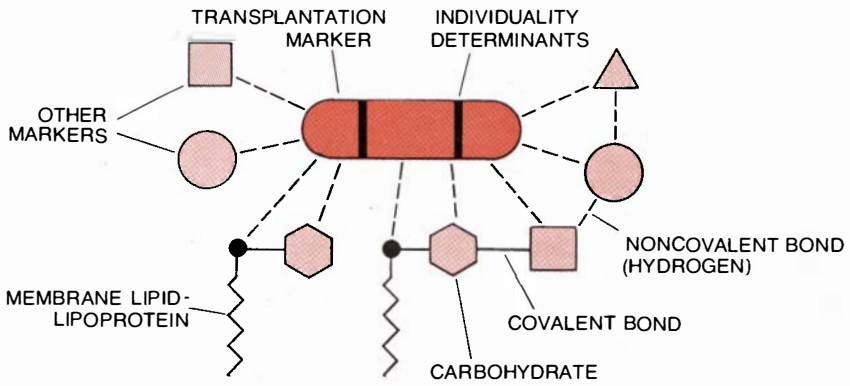
leasing markers from their membrane matrix [*see bottom illustration at right*]. One method uses proteolytic enzymes, for example papain, that randomly break the covalent bonds holding proteins together, thereby releasing an array of antigenic materials from membranes. Some of these fragments represent broken markers; others contain extraneous materials, such as carbohydrates, that apparently are not involved in the recognition of foreign cells in transplantation.

A gentler method of disruption, which leaves the markers more or less intact, involves exposing cells to low-intensity sonic energy. Noncovalent bonds are disrupted by a combination of cavitation (the rapid expansion and violent collapse of air bubbles trapped in the medium), mechanical agitation, foaming, shearing and local heating. The treatment produces a complex mixture of markers, other solubilized membrane components and soluble proteins that are released from the interior of the cell by the disintegration of its surface membrane.

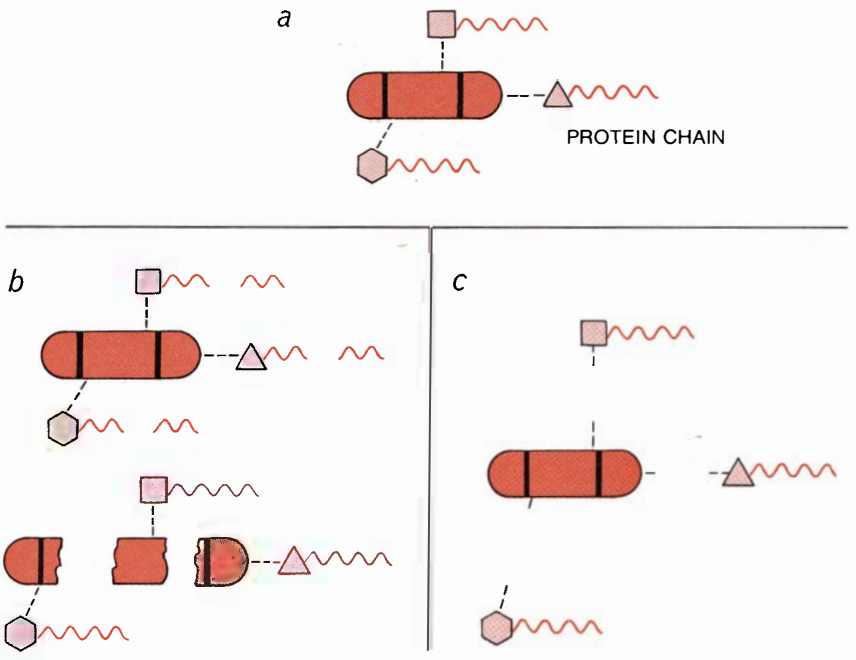
Noncovalent bonds can also be dissociated by a simple salt, potassium chloride, which acts as a weak chaotropic agent, decreasing the orderly arrangement of water molecules in the medium surrounding the cell membrane. Irving M. Klotz and his colleagues at Northwestern University have postulated that a decrease in the ordered structure of water makes it possible for hydrophobic parts of membrane-embedded proteins to become detached from their lipid environment and to become dispersed in an aqueous one. Thus by subjecting cells to a strong solution of potassium chloride for about 16 hours one can obtain a crude extract containing a variety of cell components, including intact membrane proteins in solubilized form. Since the protein markers of cell individuality represent only a minor fraction of the entire crude mixture, their extraction and purification present a challenging task.

One separation technique that has proved successful is polyacrylamide-gel electrophoresis, originally developed by Leonard Ornstein and Baruch J. Davis of the Mount Sinai School of Medicine. Its usefulness has been considerably enhanced by a computer program based on theories developed by Andreas Chrambach and David Rodbard of the National Institutes of Health. With this program we have been able to exploit subtle differences in the molecular size and net charge of various components in the crude mixture of solubilized surface materials so that the transplantation mark-

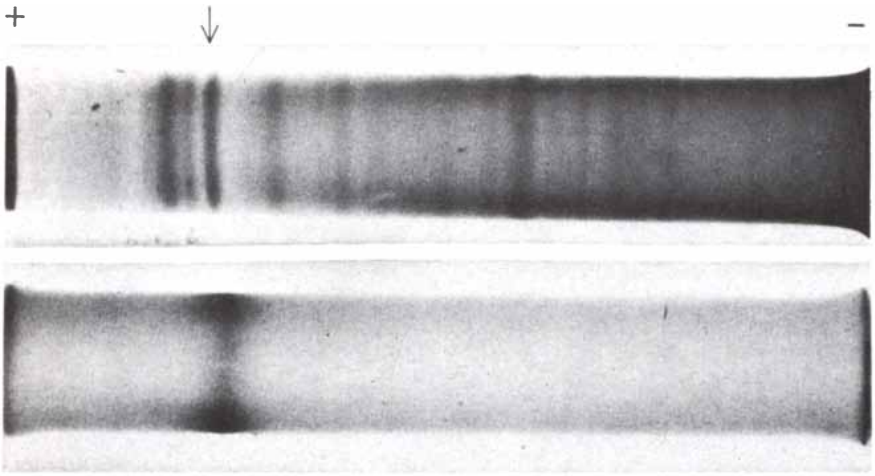




INDIVIDUALITY MARKERS that trigger a strong immunologic response in tissue transplantation are proteins (*dark color*) distributed in the surface membrane of virtually all nucleated cells. The black stripes represent determinants that vary from individual to individual. The membrane also contains other proteins that act as weaker surface markers. The matrix of the membrane is a bilayer of lipid molecules. The strong transplantation markers are bound to other membrane units (*bottom diagram*) largely through noncovalent bonds, which are more easily broken than the covalent bonds that hold molecules together.



RELEASE OF MARKERS FROM CELL MEMBRANE can be accomplished by treating cells with proteolytic enzymes, which break covalent bonds, or by using milder methods that disrupt mainly noncovalent bonds. A marker is shown before treatment in *a*. Two possible results of treating cells with proteolytic enzymes are shown in *b*. In one case (*top*) the enzyme breaks the bonds of proteins loosely attached to the marker; in the other case (*bottom*) the enzyme breaks up the marker itself. The marker remains intact (*c*) if cells are subjected to low-energy sound or to solutions of certain simple salts, such as potassium chloride.



PURIFICATION OF MARKER ANTIGENS can be achieved by subjecting a crude membrane extract to polyacrylamide-gel electrophoresis. The separation exploits differences in size and charge affecting the rate at which proteins and other cell components move through a gel medium of defined pore size in response to an electric field. The image at the top shows the electrophoretic pattern of a crude extract containing a human transplantation marker (*arrow*) known as HL-*A* antigen. At the bottom the highly purified antigen isolated from a cultured human lymphoid cell line, designated RPMI 1788, is clearly separated.

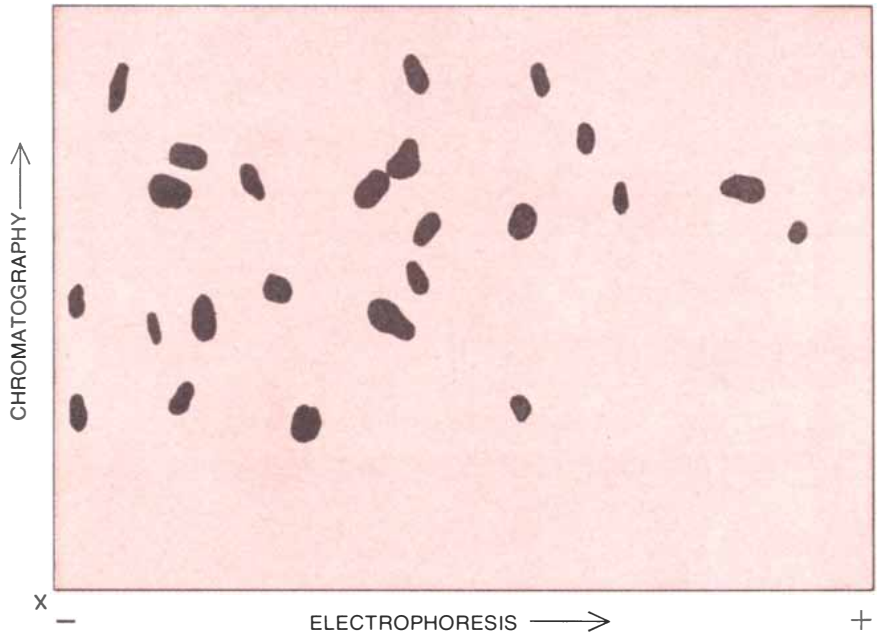
ers separate clearly. In this way the markers carried by the cells of guinea pigs and the cells of humans have been extracted as homogeneous substances with high biological activity [*see illustration above*].

The purified markers turn out to belong to a family of proteins that differ from one another in their specific amino acid composition. It was exciting to find that individuality differences heretofore undetectable except by transplantation between two strains of guinea pigs or two unrelated human subjects could be predicted by specific differences in the amino acid composition between their isolated protein surface markers. Our findings supported earlier work by Andrew A. Kandutsch of the Jackson Memorial Laboratory showing that the specificity of antigens arrayed on membrane fractions was destroyed by chemicals that attacked proteins. Thus antigenic activity resides entirely in membrane proteins and not, for example, in carbohydrates that might be associated with them. This was confirmed by Stanley Nathenson of the Albert Einstein College of Medicine in New York, who demonstrated that the carbohydrate fraction of solubilized *H*-2 antigens seems to be devoid of antigenic activity.

Transplantation markers are therefore simple proteins consisting of about 300 amino acid units strung together in a linear chain. Chemical studies show, however, that the markers of unrelated individuals have major differences in amino acid composition at five sites in the chain. As is well known from studies of the genetic code, a sequence of three bases in DNA, the genetic material, is needed to specify a particular amino acid in a protein molecule. In the case of at least four of the five variable amino acids a change in only a single base in one of the coding triplets is enough to account for the observed differences in the marker proteins.

In addition there are amino acid sequence differences involving the ordering of units in individual chains. When Michele Pellegrino in our laboratory subjected marker molecules to a proteolytic enzyme (trypsin) that attacked the covalent bonds adjacent to specific amino acids (lysine and arginine), he obtained 24 fragments, as determined by "fingerprinting" analysis [*see bottom illustration on this page*]. This is a technique in which a treated sample is placed in the corner of a square of filter paper; fragments are separated by chromatography in one direction and by electrophoresis in a direction at right angles to the first. The first separation reflects the rate at which fragments are carried across the paper by solvents; the second separation reflects the rate at which fragments move in an electric field. The resulting pattern provides a "fingerprint" of the protein.

The fingerprint of markers of two unrelated human subjects showed only six differences in the location of 24 fragments. This means that the amino acid composition of 18 of the fragments was the same in both subjects. The other six fragments exhibited amino acid differences. These findings suggest that transplantation markers are superimposed on a backbone of constant amino acid sequence but that interspersed along the



FINGERPRINT OF HUMAN ANTIGEN is produced by a technique that combines electrophoresis with chromatography. The purified protein antigen isolated from cell line RPMI 1788 is treated with a proteolytic enzyme that breaks the molecular chain wherever it contains amino acid subunits of lysine or arginine. The chain is cleaved at 23 places, yielding a total of 24 fragments. The mixture is spotted at one corner (*X*) of a square of filter paper; the fragments are separated by chromatography in one direction and by electrophoresis in another. From such studies it appears that the amino acid differences in protein markers obtained from unrelated individuals are confined to six regions of the molecule.

backbone are small regions within at least some of which the amino acid sequence is varied to establish the organism's individuality.

Since transplantation markers are simple proteins, it should be possible in time to work out the complete base sequence in the structural gene that controls their synthesis. Direct gene-mapping is not possible for other strong individuality systems in mammals, such as blood-group substances, because their antigenicity depends on carbohydrates inserted in a supporting protein molecule by enzymes.

The limited variability of transplantation markers makes them much more tractable for detailed study of individuality than the immunoglobulins, the proteins that act as antibodies in blood serum, where they exhibit recognition and inactivation functions similar to those performed by cell-surface markers. The immunoglobulin molecule consists of a "constant" region whose amino acid composition varies among different individuals of the same species. This variation, known as allotypy, is not related to the molecule's recognition function in its role as an antibody. That difference is provided by the "variable" region of the molecule; each individual has the capacity to produce a vast number of different immunoglobulins (perhaps several hundred thousand) each containing a slightly different amino acid sequence in the variable region. It is not yet known how much of the far more limited variability detected in fingerprints of transplantation markers represents allotypy, or transplantation individuality differences, and how much represents functional variability.

The human marker studied by Pellegrino is known to be the product of the HL-*A*, or human leukocyte locus *A*, gene. The HL-*A*, gene product, which may be the primary determinant of whether a human graft will be tolerated or rejected, is now available for use in clinical transplantation.

LOCATION OF INDIVIDUALITY MARKERS, the cellular proteins that help to establish each organism's identity, can be visualized by fluorescent microscopy. Target lymphocytes bearing the HL-*A*, or "strong," antigenic markers of one individual were first reacted with specific HL-*A* antibody from another individual. The resulting antigen-antibody complex on the lymphocyte surface was made visible by adding a fluorescein-labeled antibody to human gamma globulin produced in a goat. The micrograph, prepared in the authors' laboratory at the Scripps Clinic and Research Foundation by P. G. Natali, shows that the antibodies recognize and affix themselves to specific antigenic sites on the cell's surface.

The distribution of transplantation markers in the general population can be studied with the methods of "tissue-typing." Individuals can be immunized against foreign tissue in several ways: by skin grafting, by organ transplants or by the natural transplant represented by the fetus in pregnancy. Immunization produces antibody that reacts not only against the particular donor's cells but also against the cells of unrelated individuals. To perform the typing, blood serum containing antibody is taken from the host and is mixed with white blood cells taken from another person [*see illustration on next page*]. Everyone whose cells are killed by the serum shares a factor foreign to the host and thus forms a tissue type.

Tissue-typing has been intensively studied during the past decade, but it is still in a period of flux. It has been found that human hosts can respond to multiple antigens in the challenge graft and not solely those related to transplantation. Furthermore, the reaction of a given individual against a set of determinants is highly complex. The reaction depends on the individual's capacity to respond to a given amount of marker, on his previous exposure to related materials that may have shared the same determinants, to say nothing of his own genetic constitution. As a result the superimposition of the individuality of immune responses on the vast genetic variety in transplantation markers and other markers yields a bewildering array of results, restricting the present value of typing in predicting the best unrelated donor for a given patient in need, say, of a kidney transplant.

An important advance in typing has recently come out of Ferrone's work in our laboratory. Ferrone finds that when solubilized human antigens (individuality markers released from cell membranes) are administered to rabbits, the animals' immune system can recognize individual differences it cannot discern when it is confronted with whole human cells. The explanation is that intact cells carry species-specific markers that are much stronger than the markers that vary from individual to individual. The purified soluble material evidently does not possess the species markers. As a result the rabbits respond to the individuality markers rather as another human being would. Studies involving many rabbits have shown that some animals respond to only a limited number of leukocyte locus *A* markers, and in some cases to only a single marker. These studies promise to improve tissue-typing and so lead to more successful organ-matching.

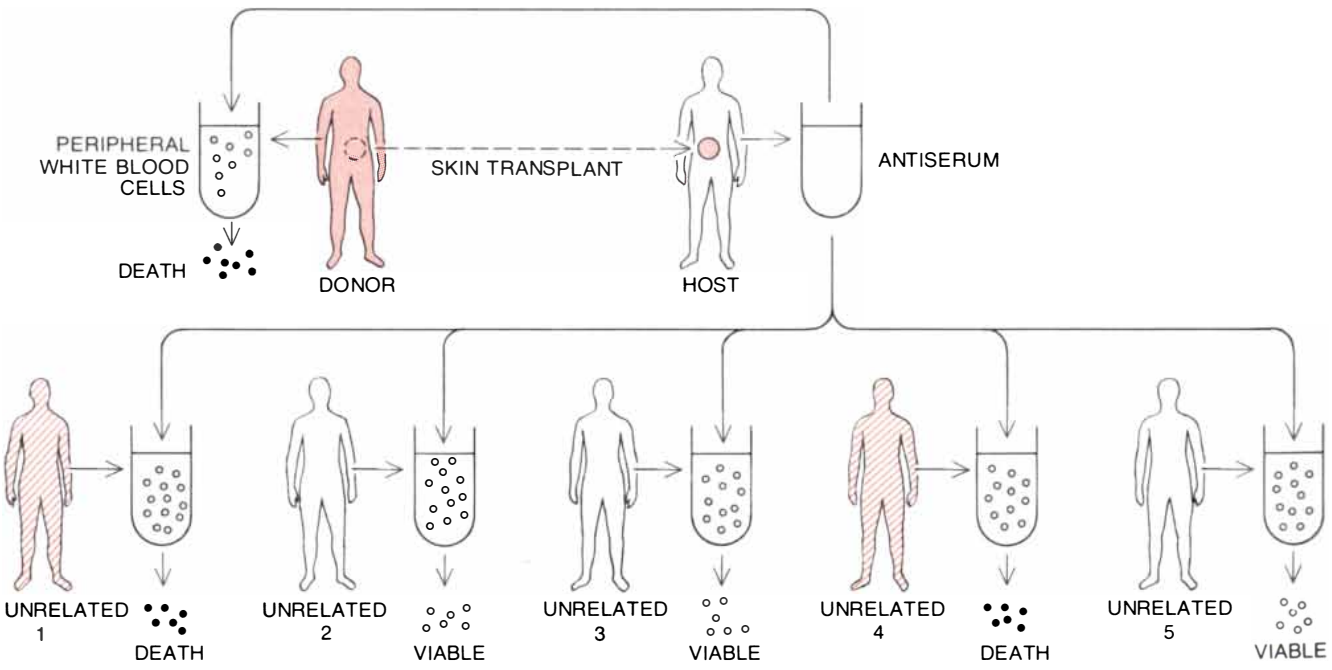
Other potential clinical applications of solubilized transplantation markers will be investigated in our transplantation unit at Northwestern University. The

markers may provide an index to measure the patient's response to grafts so that the surgeon can regulate the amount of drugs necessary to suppress the patient's immune response and thereby enable the graft to survive. Another approach, which is now being tested in animals, is to pretreat the host with purified solubilized individuality markers obtained from the prospective donor. The host normally confronts such markers on the donor's intact cell surfaces; there is evidence that his immune system may be confounded and respond less vigorously or erratically when the markers are first presented in solubilized form. Under the circumstances his body may accept the markers as his own and fail to react at all, thus exhibiting immunological tolerance. He may develop a response that actually protects the subsequent graft rather than destroying it, a response termed immunological enhancement. Or his response may be an impotent "deviant" reaction that neither destroys the graft nor protects it. Whichever the case, experiments with animals demonstrate that the survival of a graft is significantly prolonged by pretreating the host with a solubilized individuality marker. Immunotherapy with these substances may well provide a major step in controlling the rejection of tissue transplants and break the barrier that has so far limited the clinical transplantation of organs.

As we have observed, individuality markers are present on virtually all nucleated cells and hence must play an essential role in the cell's economy. Since they spread freely across cell surfaces, it seems unlikely that they are involved in the cell structure itself. Conceivably they help to regulate the permeability of the cell membrane. Another possibility is that they may assist cells of the same type to form aggregates. Thus they may facilitate the adhesion of cells in the architecture of tissues or enable cells to exchange information. These functions might be a property of the structurally constant regions of molecules that also possess localized regions of individuality. It seems reasonable to suppose natural selection would have endowed marker molecules with one or more functions in addition to providing an identification label unique for each organism.

One puzzling observation is that cells from individuals never previously exposed to foreign markers in grafts often act as if they had encountered the markers before. H. Sherwood Lawrence of the New York University School of Medicine has suggested that such individuals may have encountered the markers, or close copies of them, in molecules carried by bacteria or viruses. This idea is supported by the fact that grafts, like intracellular bacterial and viral parasites, are destroyed by a cellular immunological mechanism. It seems entirely possible that each person is characterized not only by his innate individuality markers but also by an entire menagerie of infectious agents to which he has been exposed and whose markers he carries around throughout his life. This suggests in turn that a person's own markers may either help to protect him from certain disease processes or increase his susceptibility to them. In other words, in order to attack a cell successfully a bacterium or a virus might have to play a molecular game of wits with the individuality markers and with the immune potentials of the host that stand in its way. The hypothesis is supported by the observation that certain anti-HL-*A* antibodies that block leukocyte locus *A* marker sites also interfere with the infectivity of viral agents, thus suggesting that the agent shares the determinants.

There is evidence that various diseases are associated with leukocyte locus *A* factors, indicating that individuality markers are indeed related to the inception, development and pathogenic reaction to disease. On the other hand, the host's life might be prolonged if he were fortunate enough to harbor a para-

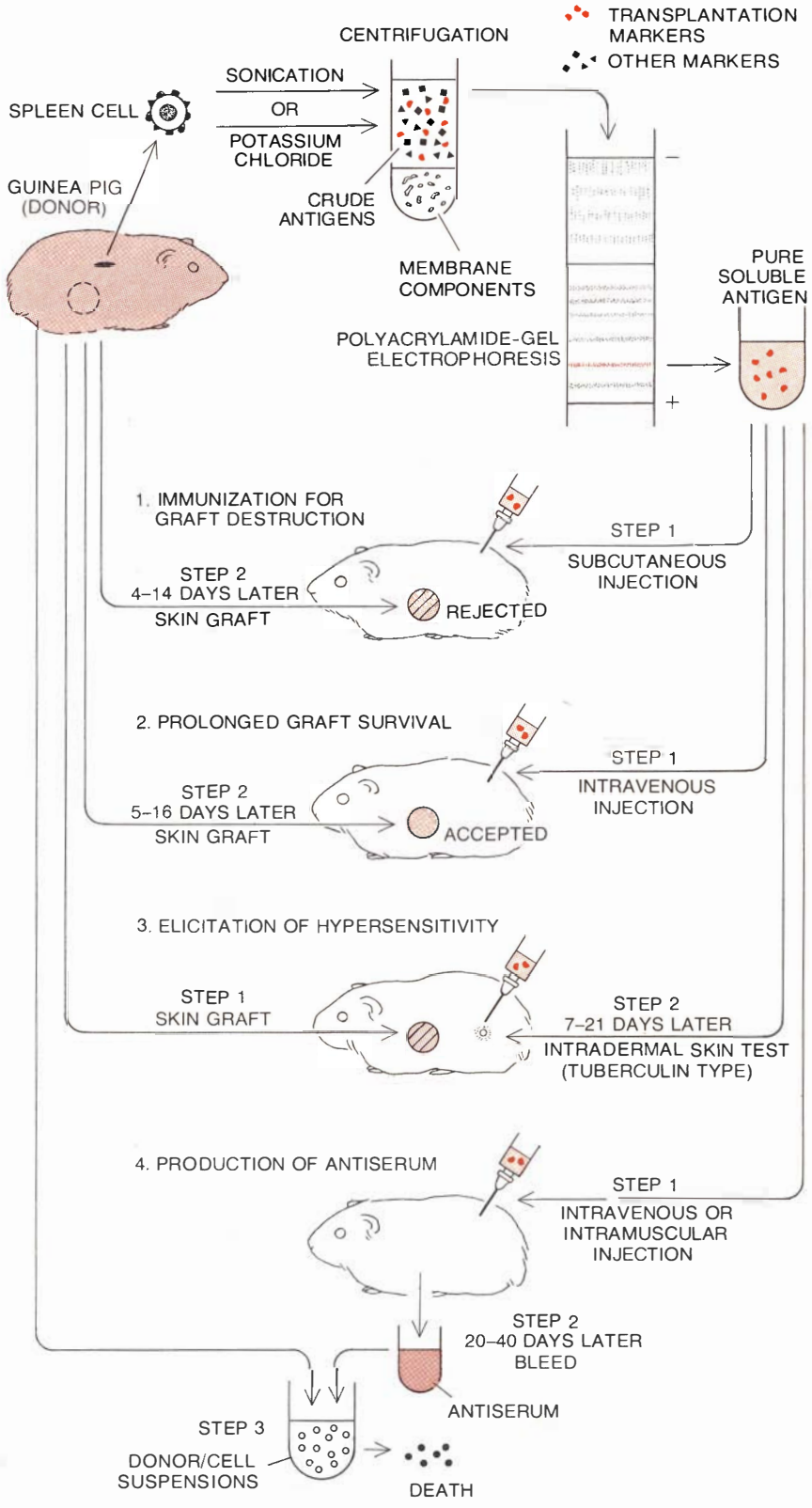


TISSUE-TYPING has been used to study the distribution of transplantation markers in the general population. An individual (*"host," top*) who has been immunized against an unrelated person (*"donor"*), for example by a skin graft, produces antiserum, a serum that contains antibody. The antiserum will destroy not only the donor's white blood cells but also the white cells of other unrelated individuals (*1, 4*). All the individuals whose cells are so destroyed are considered to belong to a common tissue type. Tissue-typing has been helpful in selecting related donors for organ transplants but is only modestly useful with unrelated individuals.

site that supplied markers he lacked. For example, it has been reported that leukemia and Burkitt's lymphoma have regressed after a patient had contracted measles. One might say that the perfect parasite is one that succeeds in lengthening the life of its host—but then should it still be called a parasite?

The individuality-marker system may protect the species against viral infections in another way. When some viruses that contain RNA rather than DNA as their genetic material infect a cell, they form buds and carry a portion of the cell's surface membrane along with them. If the cell's individuality markers thereby end up in the coat of the virus, they could elicit a "graft" rejection when the virus attacks another unrelated individual of the same species as the original host. The possibility that transplantation individuality markers can determine the host's response to viruses is potentially important in clinical diagnosis, prophylaxis and immunotherapy.

Recently Burnet has elaborated on a suggestion made earlier by Lewis Thomas of the Yale University School of Medicine to construct a comprehensive hypothesis of an immune surveillance system. Early in evolution primitive organisms developed a class of wandering cells capable of recognizing and destroying foreign cells or parasites. This mechanism, Burnet points out, would be able to act not only against agents of external origin but also against mutated cells of the host organism that exhibited new individuality markers. This would tend to make the individual resistant to new growths, that is, tumors and cancers in general. Functional surface markers on the host's own cells would serve as local recognition sites for the wandering surveillance cells; in addition markers released from cell membranes in the normal course of cell repair and replacement would telegraph subcellular messages to draining lymph nodes, keeping them informed of their constituency. Natural selection would favor such a system because it enhances the viability of the organism. Burnet suggests that the aging process involves a depletion in the number and vigilance of the "wanderers," allowing such abnormalities as cancer and immune responses to the host's own tissue to develop. Therefore markers of biological individuality provide an excellent tool for probing the most intimate functions of the cell. Knowledge so obtained cannot fail to be useful in meeting the challenges of human transplantation and disease.



RESPONSE TO SOLUBLE INDIVIDUALITY MARKERS can take a variety of forms. Here pure markers, or antigens, obtained from guinea pig spleen cells are administered to four unrelated individuals. Subcutaneous injection of the antigens (*1*) will normally immunize an animal against a subsequent skin graft from the animal that supplied the antigens. On the other hand, an intravenous injection of antigens (*2*) will often prolong the survival of a skin graft made from five to 16 days later. One can tell if an animal is hypersensitive to the donor's tissue (*3*) by applying antigens from the donor in the form of an intradermal injection from one to three weeks after the skin graft. Finally, one can produce antiserum by injecting the antigens subcutaneously in an unrelated host (*4*). After a few weeks the host's serum will contain enough antibodies to destroy the donor's white cells.

ORGANIC MATTER IN METEORITES

The meteorites known as carbonaceous chondrites contain organic compounds. These compounds have now been subjected to detailed analyses which indicate that they are not of biological origin

by James G. Lawless, Clair E. Folsome and Keith A. Kvenvolden

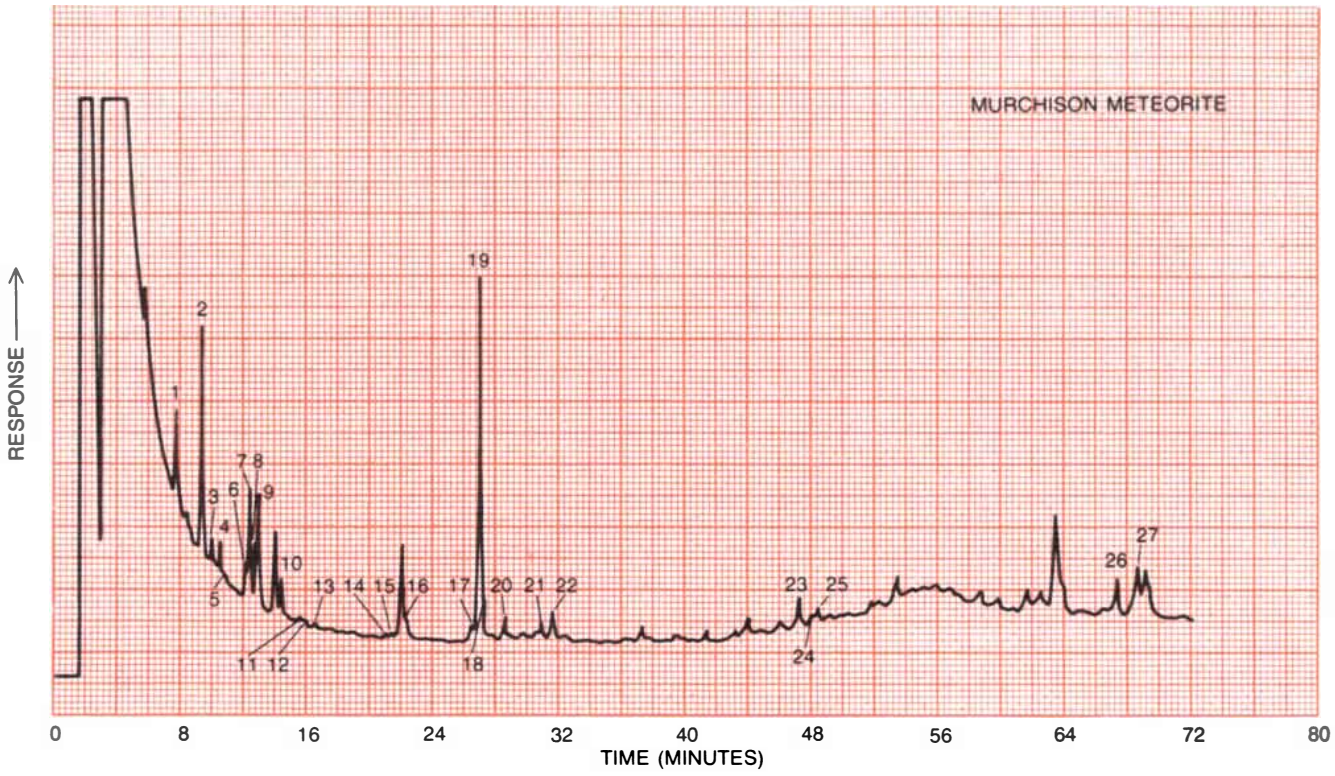
On the morning of Sunday, September 28, 1969, an object exploded over the town of Murchison in Australia and showered incandescent fragments over an area seven miles long and two miles wide. Although few people actually saw the falling fragments, many heard the noise; it lasted almost a minute and sounded like thunder or a series of sonic booms. With this fanfare the Murchison meteorite arrived on the earth. At first reports some 4.5 kilograms of fragments were found, but much more material actually fell, since at least 82.7 kilograms are now in public collections. On chemical analysis the meteorite was classified as a Type II carbonaceous chondrite, joining about 15 other meteorites that are in this category.

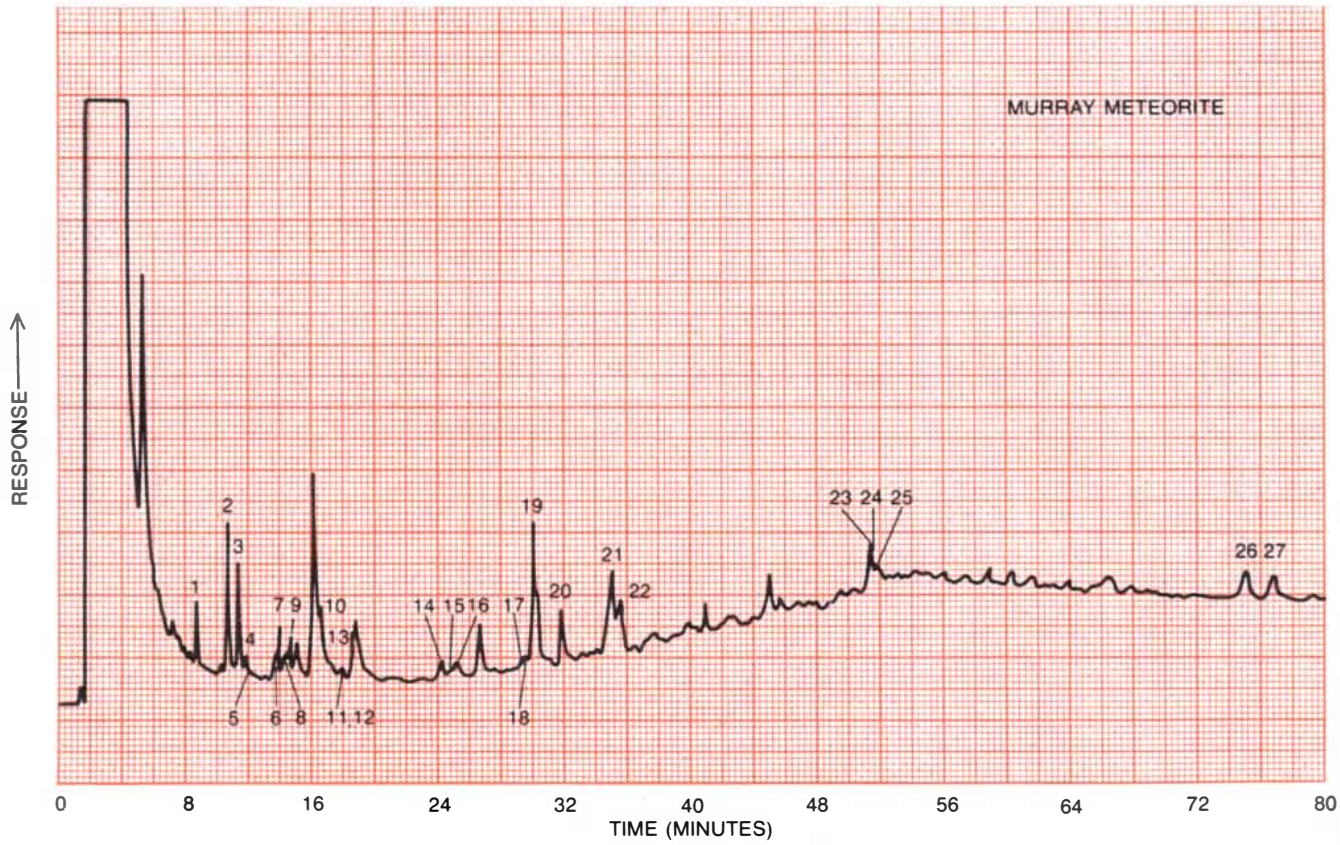
Carbonaceous chondrites are quite unlike other meteorites. They are black objects, easily crumbled or pulverized, and have an appreciable amount of carbonaceous material other than free carbon. They are called chondrites because they contain chondrules: small round bodies consisting of the magnesium-iron silicates olivine and pyroxene. Detailed chemical studies have recently shown that the Murchison meteorite contains a number of organic (carbon) compounds never encountered before in meteorites. The organic compounds in the Murchison meteorite seem to have been synthesized entirely by nonbiological processes. These findings have inspired a renewed interest in the organic constituents of meteorites and have provided new clues to the origin and significance of organic substances in the solar system.

The fact that meteorites contain organic substances has been known for more than a century. As long ago as 1834 Jöns Jakob Berzelius of Sweden extracted complex organic substances from a meteorite that had fallen near Alès in southern France. He wondered about the significance of his discovery with regard to the possibility of extraterrestrial life. Even today, after more than 140 years of work, investigators still ponder the origins of the organic material in carbonaceous chondrites. Following Berzelius' discovery a number of chemists applied increasingly more advanced techniques to the study of carbon compounds in meteorites. Brian Mason, who was then working at the American Museum of Natural History, summarized the investigations up to 1962 in this magazine [see "Organic Matter from Space," by Brian Mason; SCIENTIFIC AMERICAN, March, 1963]. By then it had been clearly shown that meteorites contain hydrocarbons, compounds consisting solely of carbon and hydrogen.

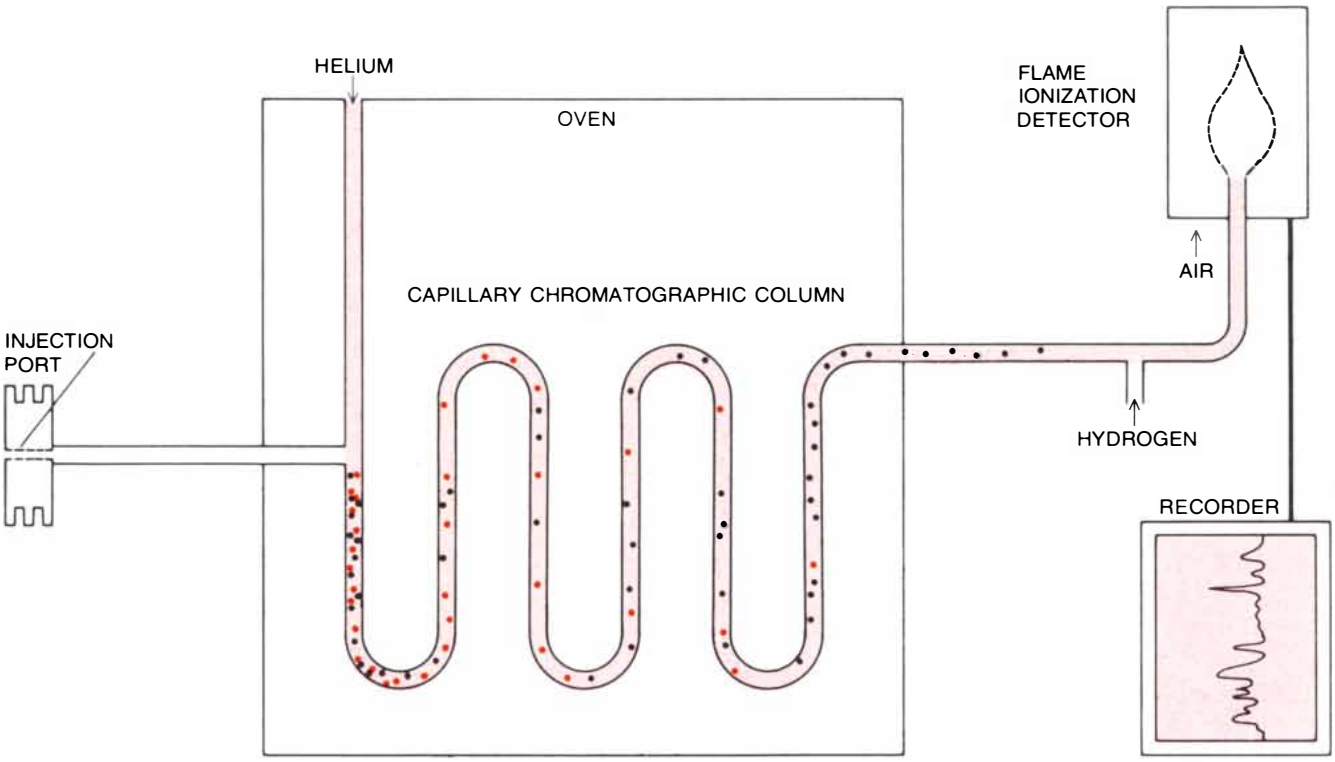
During the next eight years many investigators directed their research toward finding different classes of organic compounds in meteorites and elucidating the structure of the individual organic molecules. Not only did they study hydrocarbons in more detail; they also reported such compounds as carboxylic acids, alcohols, phenols, sugars, amino acids, purines, pyrimidines and porphyrins. All these compounds are associated with life processes on the earth. It is certain that much of the organic material in carbonaceous chondrites is of extraterrestrial origin, but it is also well recognized that some of the individual compounds could very well be biological contaminants that have entered the meteorites from the air or the ground during the time that elapsed between their fall to the earth and their recovery. The problem of determining

MURCHISON METEORITE, a Type II carbonaceous chondrite, fell near Murchison, 85 miles north of Melbourne, Australia, in 1969. This is only one of many fragments recovered.

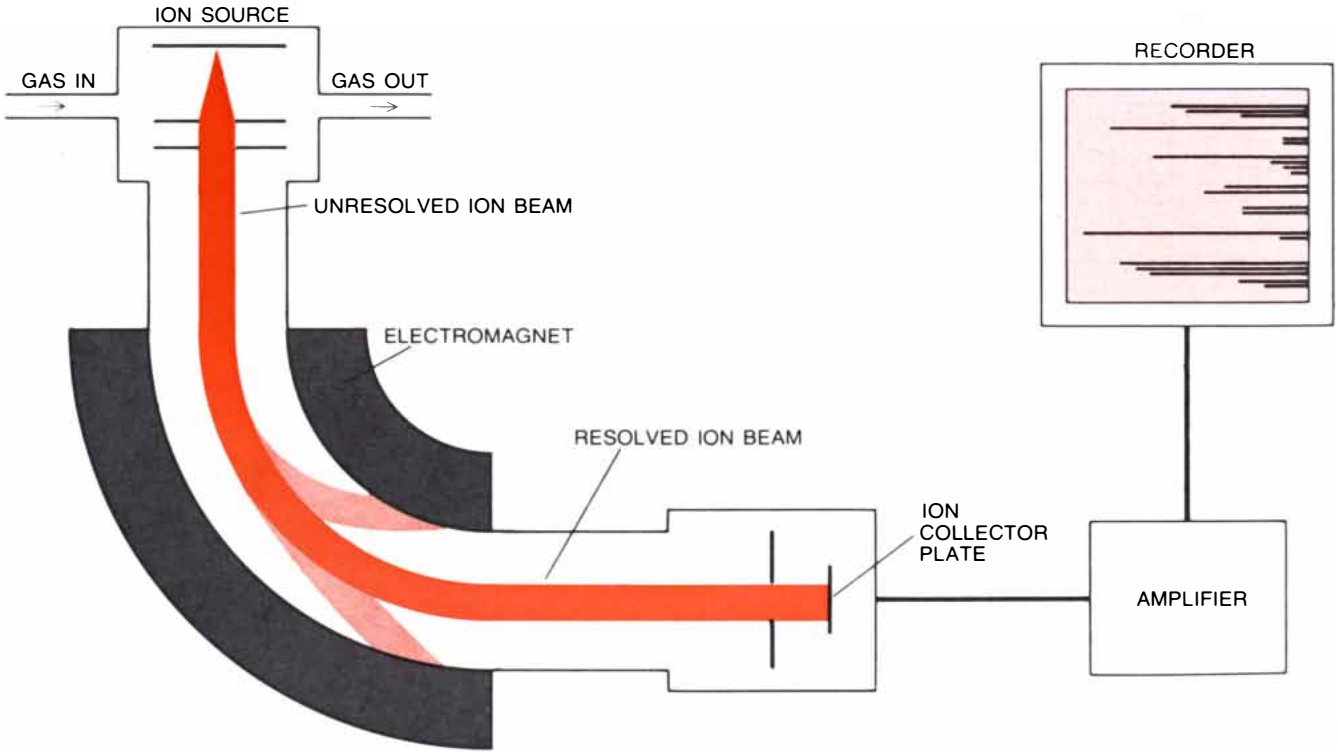




GAS CHROMATOGRAMS reveal that the Murchison meteorite (*top*) and the Murray meteorite (*bottom*) have a quite similar chemical composition. At least 18 amino acids are present in both; nine could be separated into *D* and *L* forms, present in equal abundances for each amino acid. The compounds identified are (*1*) isovaline, (*2*) α-aminoisobutyric acid, (*3, 4*) *D*- and *L*-valine, (*5*) *N*-methylalanine, (*6, 8*) *D*- and *L*-α-aminobutyric acid, (*7, 9*) *D*- and *L*-α-alanine, (*10*) *N*-methylglycine, (*11*) *N*-ethylglycine, (*12, 13*) *D*- and *L*-norvaline, (*14, 15*) *D*- and *L*-β-aminoisobutyric acid, (*16*) β-aminobutyric acid, (*17, 18*) *D*- and *L*-pipecolic acid, (*19*) glycine, (*20*) β-alanine, (*21, 22*) *D*- and *L*-proline, (*23*) γ-aminobutyric acid, (*24, 25*) *D*- and *L*-aspartic acid, (*26, 27*) *D*- and *L*-glutamic acid.



GAS CHROMATOGRAPH is used for a preliminary analysis of the organic compounds found in the Murchison meteorite. The sample of compounds is volatilized into a carrier gas, in this case a stream of helium. The helium and compounds pass into a capillary column .02 inch in diameter and 150 feet long where the various components of the sample migrate at different speeds. They enter a flame ionization detector, which burns the compounds and ionizes the molecules, increasing the electrical conductivity across the flame with respect to the conductivity of the helium. The detector records this difference on a strip chart as a gas chromatogram.



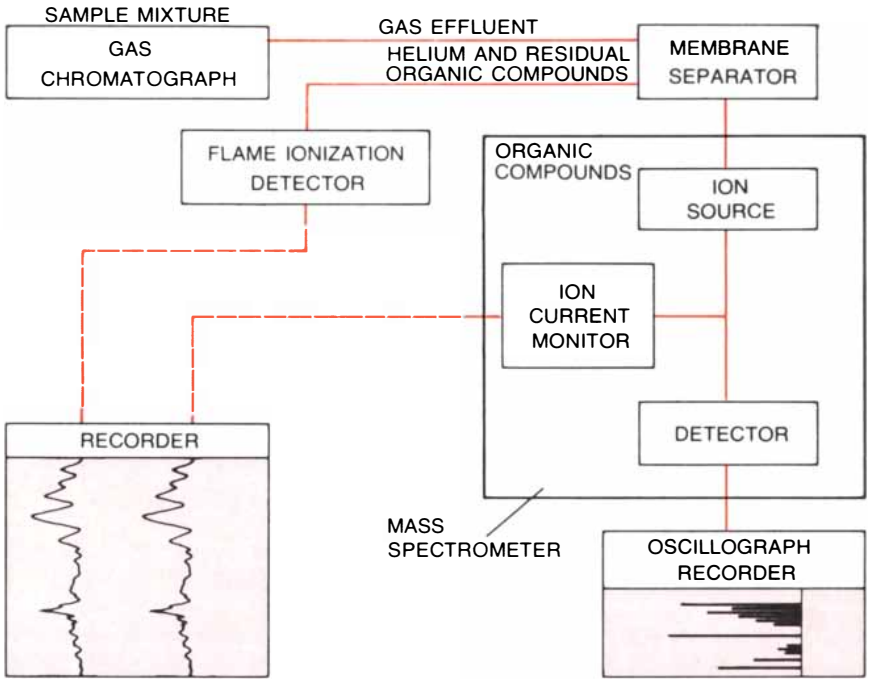
MASS SPECTROMETER analyzes the masses of atoms or molecules in a sample of compounds in the Murchison meteorite. Molecules of the meteorite sample are introduced into an evacuated chamber where they are bombarded by a beam of electrons and are ionized or broken into charged fragments. The ion beam (*color*) is bent by an electromagnet. The ions that reach the collector plate signal their presence by a minute electric current. Adjusting the magnetic field that bends their path brings ions of the desired weight to the collector. The distribution of the positive ions with respect to their mass is recorded sequentially as a mass spectrum.

the origin of organic substances in meteorites can be rigorously attacked only if the extraterrestrial and terrestrial contributions are recognized and separated. Recognition is difficult because many molecules are structurally the same whether they are generated extraterrestrially in abiotic (nonbiological) processes or are synthesized by biological systems on the earth.

The Murchison meteorite thus has a particular significance. Since a number of fragments were collected soon after impact, they appear to have a minimal amount of terrestrial contamination. And since the organic compounds in the meteorite seem to be almost entirely nonbiological, it is possible for the first time to begin to define the characteristics of organic compounds that have resulted from abiotic extraterrestrial synthesis. The synthesis may be part of a grand-scale sequence of events in the chemical evolution of the universe.

The Oparin-Haldane hypothesis of chemical evolution, put forward by A. I. Oparin of the U.S.S.R. and further developed by J. B. S. Haldane of Britain, postulates that complex organic molecules were formed from simple precursors before the appearance of life. Various forms of energy—ultraviolet radiation from the sun, heat from volcanoes, the electrical discharges of lightning, ionizing radiation from radioactive nuclei—acted on the primitive atmosphere to produce on the earth the complex organic molecules that eventually led to the first life forms. The hypothesis of chemical evolution and the observation of chemical processes in the cosmos suggest that what took place on the earth more than three billion years ago may well have taken place or may currently be taking place on innumerable planets in the universe.

In any investigation of chemical evolution a number of assumptions are made. First, it is assumed that compounds that are of biological importance today were also important at the time life originated. This assumption seems reasonable because there are certain molecules that are characteristic of life from its simplest forms to its most complex. Second, it is assumed that there is nothing unique about our sun and solar system. There are more than 10^{20} stars in the universe, and conservative estimates suggest that there are more than 100 million planetary systems in the universe capable of supporting life.



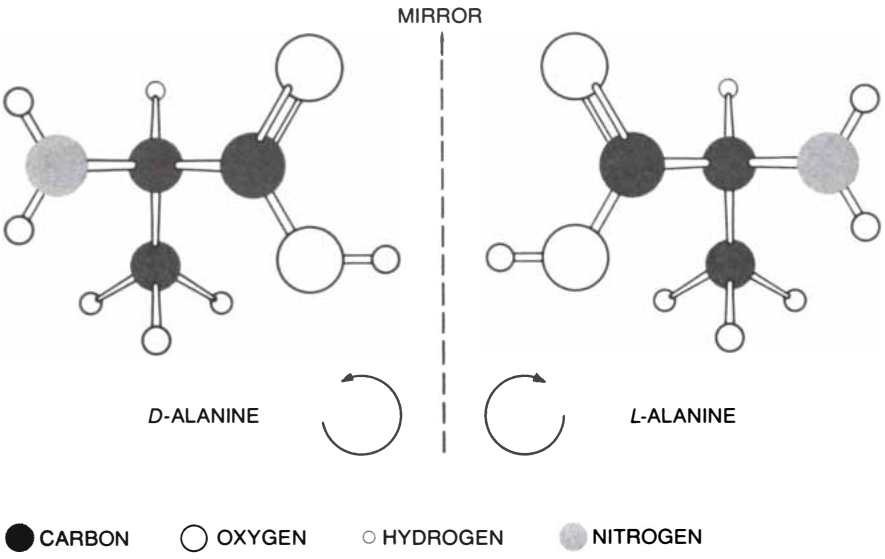
COMBINING GAS CHROMATOGRAPHY AND MASS SPECTROSCOPY allows positive identification of the meteorite amino acids to be made more readily than by either technique alone. The effluent of the gas chromatograph passes into a membrane separator, where most of each organic compound is fed into the mass spectrometer. The helium and residual organic compounds go to a flame ionization detector and produce a gas chromatogram. The molecules that enter the spectrometer are ionized; the ion current produced is directly proportional to the concentration of molecules at any instant. This current is monitored and generates a record identical with the one from the flame ionization detector. The positive ions produced are separated according to their ratio of mass to charge. The resulting mass spectrum is generally unique to the species of molecules analyzed.

Numerous laboratory experiments have been conducted in an effort to substantiate the hypothesis of chemical evolution. A classic experiment conducted by Stanley Miller in 1953 showed that molecules of biological importance could be formed by the discharge of an electric spark through a mixture of methane, ammonia and water. Many other investigators have amassed a convincing body of evidence that similar interactions in the primitive atmosphere of the earth (or the atmospheres of other planets) are capable of producing molecules of biological importance. A number of these molecules are the same ones that are characteristic of life today. To the best of our knowledge the laws of chemistry and physics are universal, and one would expect these compounds to be produced whenever the appropriate conditions exist. Since investigations of other solar systems are not feasible at present, we are confined to examining extraterrestrial bodies within our own solar system. The most readily available extraterrestrial materials are meteorites.

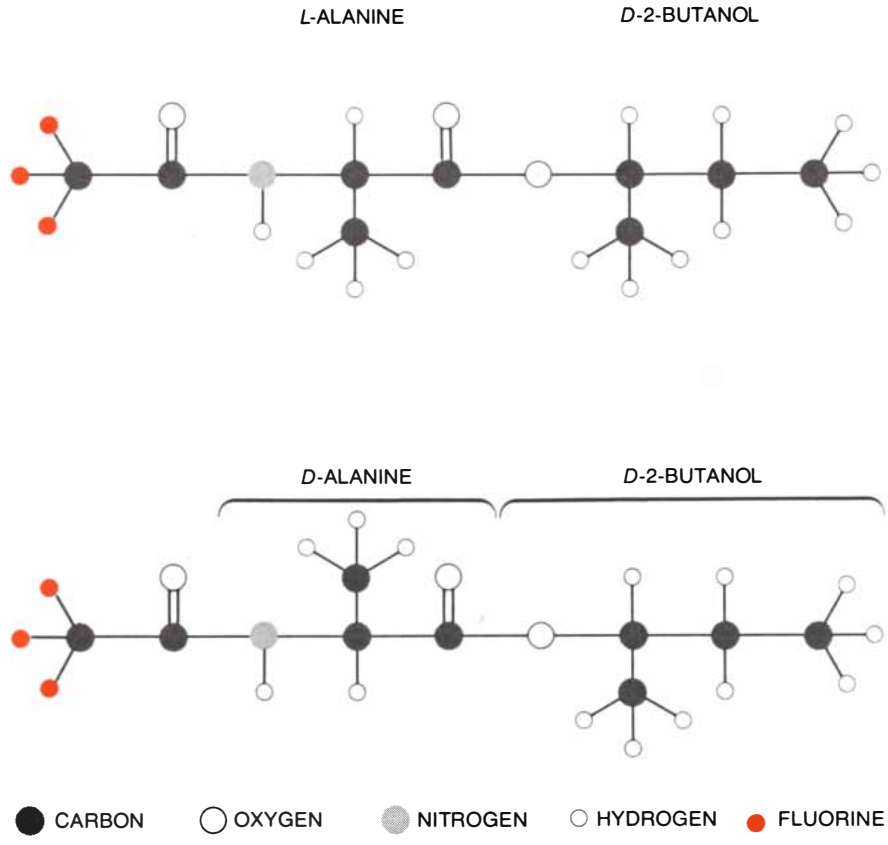
The greatest amount of work in the field of chemical evolution has been done on the synthesis of amino acids. It is therefore not surprising that this class of compounds was one of the first studied in the investigation of the organic matter indigenous to meteorites. Amino acids are relatively simple organic compounds composed of some of the most abundant reactive elements in the solar system: hydrogen, carbon, oxygen, nitrogen and sometimes sulfur and the halogens (chlorine, bromine, iodine and so on). These elements can be joined in an almost infinite variety of structures, but to be called an amino acid a structure must possess at least one amino group (NH_2) and at least one carboxylic acid group (COOH). Even with this constraint the number of theoretically possible amino acids is almost infinite.

Amino acids are of special interest because of the important roles they play in living organisms. They are the building blocks of life in that they are the monomers from which the polymers we call proteins are constructed. Proteins are found in all living cells; they hold the organism together and make its chemical processes go. The only other compounds that are as important in living systems are the nucleic acids, which are the repository of genetic information and direct the synthesis of proteins.

Some 20 different amino acids are found in proteins. All of them have a common structural feature: an amino group and a carboxylic acid group are



MIRROR-IMAGE MOLECULES, such as the amino acid *D*-alanine (*left*) and *L*-alanine (*right*), cannot be superimposed and are called enantiomers. The letters *D* and *L* designate the configuration of the four different chemical groups around the central carbon atom of each enantiomer. In this illustration *D*-alanine has the amino group (NH_2) to the left of the carbon atom and the carboxylic acid group (COOH) to the right; the reverse is true for *L*-alanine. Arrows (*bottom*) indicate that *D*-alanine rotates plane-polarized light to the left and *L*-alanine rotates this light the same amount to the right; otherwise the enantiomers have identical chemical properties. (The direction of rotation of light is not specified by the letter notations *D* and *L*, which indicate structural configuration only.)



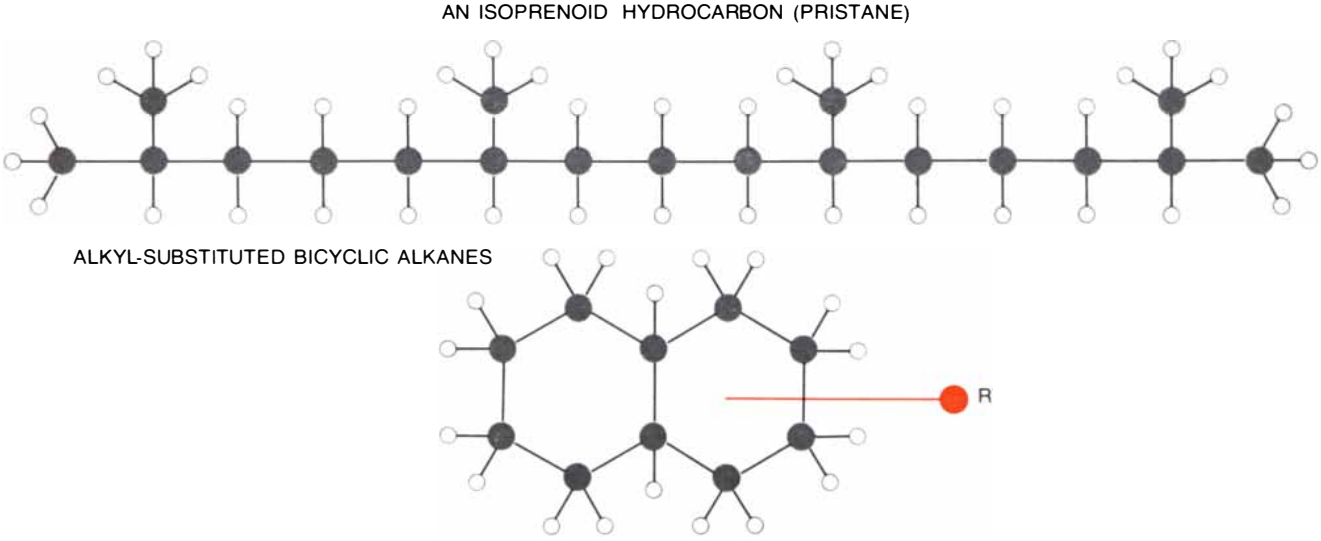
VOLATILE DERIVATIVES of alanine and other amino acids are formed by reacting the amino acids with an optically active alcohol such as *D*-2-butanol. The alcohol combines with *L*-alanine (*top*) and with *D*-alanine (*bottom*) to produce derivatives that are not mirror images, have different chemical properties and can be separated by gas chromatography.

attached to the same carbon atom. Such amino acids are known as alpha amino carboxylic acids. Amino acids other than those that make up proteins are also present in organisms, usually in low concentration. Some 150 such nonprotein amino acids have been found as individual units or as components of relatively small molecules such as peptides.

Every protein amino acid except one (glycine) has at least one asymmetric carbon atom to which four different chemical groups are attached. (Glycine has only an extra hydrogen atom.) Studies of the three-dimensional structure of protein amino acids reveal they all have the same configuration around the carbon atom bearing the amino group and the carboxylic acid group. The configuration is the same as it is in *L*-glyceraldehyde, a standard reference compound, so that these amino acids are said to have an *L* configuration. The nonsuperimposable mirror image of an *L* amino acid is a *D* amino acid. Molecules that are mirror images of each other are known as enantiomers. Such substances consist of exactly the same elements in the same quantities and have identical properties except for the direction in which they rotate plane-polarized light [*see top illustration at left*].

A solution of an amino acid containing more of one enantiomer than the other will rotate the plane of polarized light, and the mixture is said to be optically active. A mixture with equal proportions of each enantiomer is called racemic; it is optically inactive because the *L* and *D* amino acids cancel their rotation of polarized light. Individual asymmetric amino acid units in the proteins of organisms are in the *L* configuration. Each of these amino acids is optically active.

At the Ames Research Center of the National Aeronautics and Space Administration we, together with Cyril Ponnamperuma, Etta Peterson and Jose Flores, have found that the amino acids in the Murchison meteorite are optically inactive. This finding was based not on the failure of these amino acids to rotate the plane of polarized light but on another kind of analysis showing that each enantiomer of a given amino acid was present in roughly equal proportions. The ratios of the enantiomers were determined by means of a relatively new technique incorporating gas chromatography. In this technique a mixture of compounds is volatilized into a stream of helium. The compounds pass into a capillary column .02 inch in diameter and 150 feet long, where they dissolve in



ISOPRENOID HYDROCARBONS, long linear molecules such as pristane (*top*), have been found in nine carbonaceous chondrites, although none have yet been found in the Murchison meteorite. Dominant hydrocarbons in the Murchison meteorite are a complex mixture of bicyclic, or double-ring, compounds (*bottom*) with alkyl (*R*) substituents at undetermined positions on the rings.

a liquid with a high boiling point that has been coated on the inside of the column. The rate at which each compound dissolves is a function of its physical properties, which causes the compounds to separate as they proceed through the column. As the separated compounds emerge they pass into a detector, where they are ionized in a hydrogen flame. The ionization causes a change in current across the detector. This change is amplified and used to produce a chart called a gas chromatogram. Peaks on the gas chromatogram correspond to individual compounds, and the height of each peak corresponds to the relative abundance of the compound.

Amino acids are not volatile enough for separation by gas chromatography; they must first be converted into more volatile derivatives. Moreover, amino acid enantiomers can only be separated by adding another asymmetric carbon atom to both molecules. This addition is of course necessary because the unmodified enantiomers have identical chemical properties except for the direction in which they rotate plane-polarized light, and compounds with identical physical properties cannot be separated by gas chromatography because they dissolve equally well in the liquid on the inside of the capillary column. When the enantiomers (*D* and *L* configurations) of a given amino acid are reacted with an optically active alcohol (in either the *D* or the *L* configuration), the result is two derivatives, one from each enantiomer. Called diastereoisomers, these derivatives are not mirror images of each other. They have different chemical properties, including different solubilities in the capillary-column liquid, and can thus be separated by gas chromatography.

The amino acids in the Murchison meteorite were converted into volatile diastereoisomeric derivatives by reaction with the optically active alcohol *D*-2-butanol and with trifluoroacetic anhydride. The resulting derivatives were separated by gas chromatography. The chromatographic peaks resulting from each enantiomeric pair of amino acids were equal in height. Thus the asymmetric amino acids in the meteorite were present as a racemic mixture and would not be optically active.

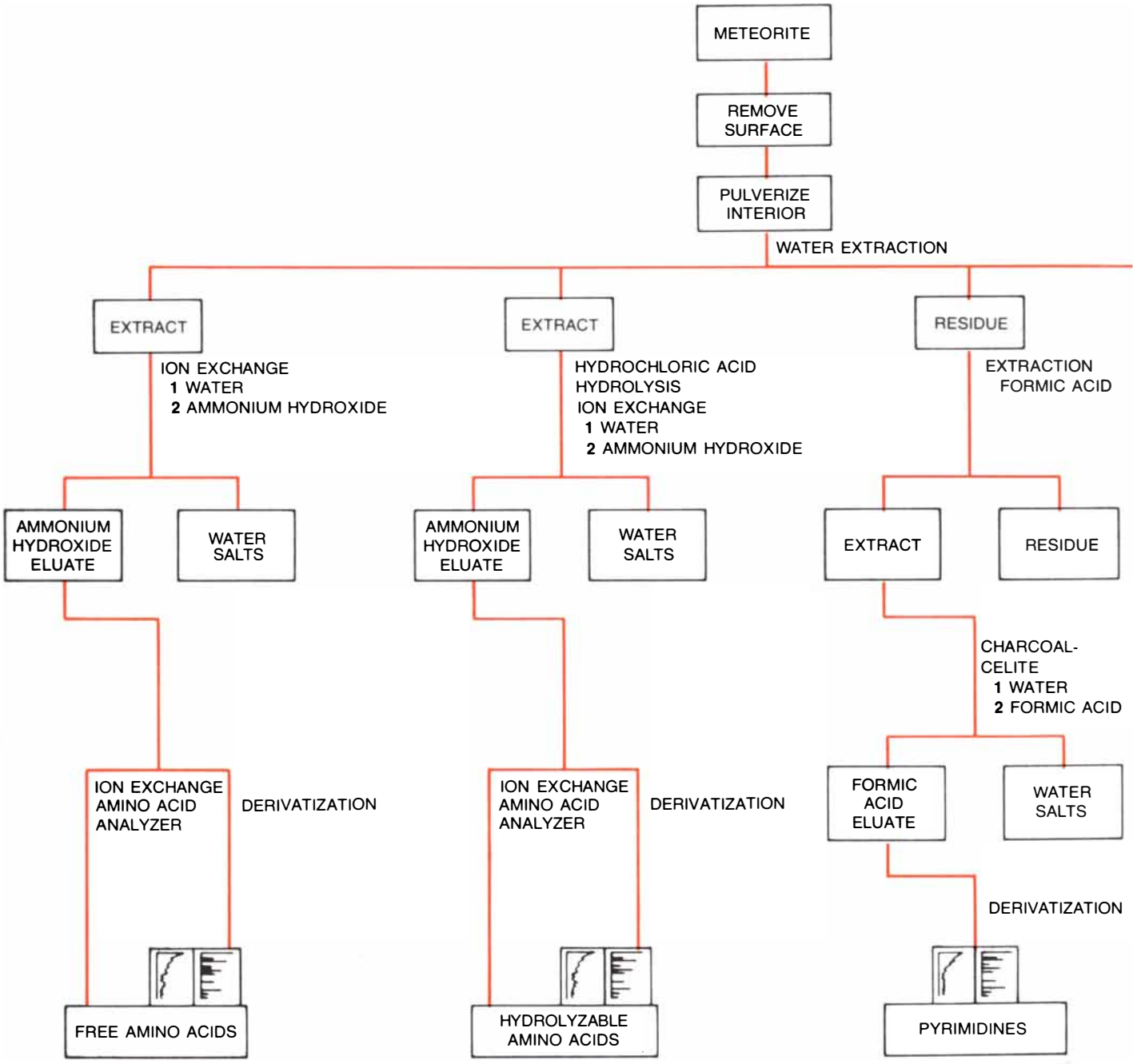
Although gas chromatography can separate amino acids and measure their amounts, it cannot readily and positively identify them. Such identifications can be made, however, by a combination of gas chromatography and mass spectroscopy. Instead of passing into the flame ionization detector the compounds are introduced into the source of the mass spectrometer; there the molecules are bombarded by a beam of electrons. The distribution of positive ions produced by this bombardment is recorded sequentially with respect to mass on a chart called a mass spectrum. Since the mass spectrum of a compound is generally unique, the spectrum of an unknown compound can be compared with spectra of standard compounds until a match is found. Once a given compound has been identified by its mass spectrum, the gas-chromatographic analysis is repeated with the addition of a small amount of the corresponding standard amino acid. The coincidence of the gas-chromatographic peaks resulting from this combination chromatography provides further evidence for the identity of the isolated amino acid.

The techniques of gas chromatography and combined gas chromatography and mass spectrometry have thus been applied to identify each amino acid and the concentration of each enantiomer in a water extract of material from the Murchison meteorite. Although there are a number of gas-chromatographic peaks that we have not yet identified, mass spectrometry indicates that most of them represent amino acids. Their absolute identification is still difficult because we do not have standard amino acids for all the molecular forms. We have nonetheless identified the major components of the mixture. All the amino acids that have asymmetric carbon atoms and whose diastereoisomeric derivatives could be separated by gas chromatography consist of nearly equal amounts of the *D* and *L* enantiomers. Of the amino acids identified, six (valine, alanine, glycine, proline, aspartic acid and glutamic acid) occur frequently in proteins. In proteins, however, they are all in the *L* configuration. The remaining 12 amino acids identified are nonprotein amino acids. They are present in low concentration in biological materials and generally as the *L* enantiomer.

Since the first report of amino acids in the Murchison meteorite, work by Juan Oró's group at the University of Houston and by Carleton B. Moore and John Cronin at Arizona State University has substantiated these results. We have also examined the Murray meteorite, a Type II carbonaceous chondrite that fell near Murray, Ky., in 1950. As Eugene

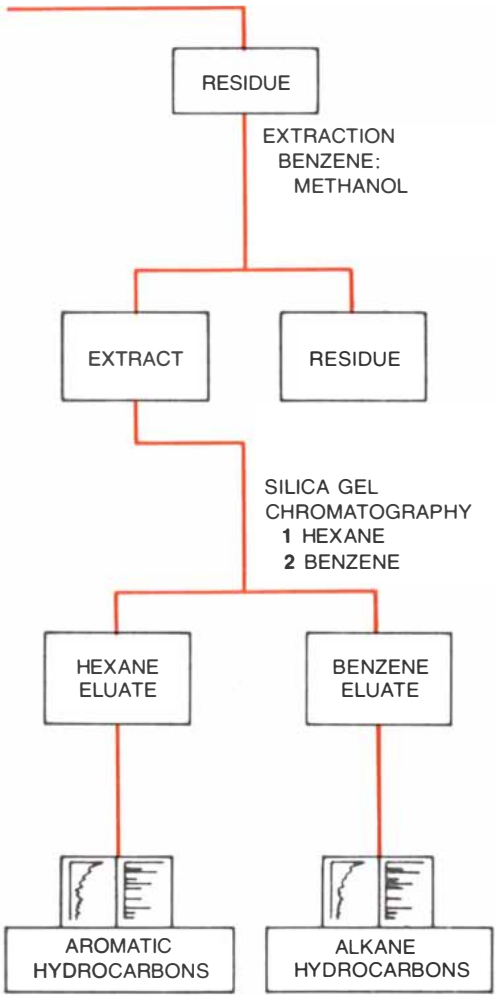
Jarosewich of the Smithsonian Institution has pointed out, the Murchison and Murray meteorites are chemically the most similar of any two Type II carbonaceous chondrites recovered and analyzed so far. Our investigation of the amino acids extracted from the Murray meteorite revealed that their distribution is similar to the one found in the Murchison meteorite. Our recent analysis of the Orgueil meteorite, a Type I carbonaceous chondrite that fell in France in 1864, shows both amino acids that are the result of terrestrial contamination and amino acids that appear to be indigenous to the meteorite.

Now that these meteoritic amino acids have been identified and their *D* and *L* characteristics established, we can inquire about their origin. Paul B. Hamilton of the Alfred I. du Pont Institute has said of earlier analyses of amino acids in other meteorites: "What appears to be the pitter-patter of heavenly feet is probably instead the print of an earthly thumb." The amino acids identified in the Murchison and Murray meteorites and the distribution of their *D* and *L* forms argue against that possibility. At this point the question arises: Were extraterrestrial life processes responsible for the amino acids in the two meteorites? It is conceivable that the amino acids once were in one stereoisomeric form and then were racemized: converted into a mixture of enantiomers by the interconversion of *D* and *L* forms over a long period. Although this argument cannot be entirely discounted, it would be difficult to explain the assortment of amino acids detected by any analogy with terrestrial life. The amino acids in



IDENTIFICATION OF ORGANIC COMPOUNDS in meteorites begins with the careful selection of a stone exhibiting the least amount of weathering and exterior contamination. The surface is chipped away and the resulting interior piece is pulverized. The sample is then extracted with various solvents to remove different classes of compounds. The extracts are analyzed by gas chromatography and combined gas chromatography–mass spectrometry. In certain cases derivatives are prepared to make the compounds

terrestrial organisms are all alpha amino acids. The meteorite contains all possible isomers of amino acids with two and three carbon atoms, and all but two of the isomers of amino acids with four carbon atoms. This distribution suggests that the amino acids were synthesized at random rather than synthesized selectively by organisms as we know them. Some of these same amino acids have been found in chemical-evolution experiments in the laboratory. Thus the amino acids in the Murchison and Murray meteorites seem to be the product of a naturally occurring chemical evolution. Because of the environment of their formation, however, the process stopped rather than proceeding on to life.



more suitable for these techniques. For the amino acids, ion-exchange chromatography is also used for removing salts and for separating and analyzing all these compounds.

Amino acids are only one of many classes of organic compounds present in meteorites. We have dwelt on this class of compounds because amino acids are ubiquitous in terrestrial living systems and it is possible that these same compounds may indicate the existence of extraterrestrial life. The nitrogenous bases classified as purines and pyrimidines form another important class of molecules in living systems. Five of these compounds are important building blocks of the nucleic acids DNA and RNA. In 1964 Ryoichi Hayatsu of the University of Chicago found two purines—adenine and guanine—in a sample of Orgueil meteorite. His discovery, however, has not yet been confirmed by other investigators.

When the Murchison meteorite became available and when indigenous amino acids were isolated from it, we began a new search for purines and pyrimidines. As with the amino acids, both gas chromatography and gas chromatography combined with mass spectrometry were applied to extracts of the meteorite that had been obtained with a number of solvents. To make volatile derivatives of the compounds for gas chromatography, reactive hydrogens were replaced with trimethylsilyl groups: $(CH_3)_3Si$–. A typical chromatogram of the compounds extracted from the Murchison meteorite with formic acid shows several large peaks. Two major classes of heterocyclic compounds (compounds with ring structures incorporating both carbon and nitrogen) were found; other tentatively identified compounds were present in small concentrations. The surprise was that none of the five common purines or pyrimidines present in biological organisms could be found, not even the simple compounds adenine and guanine. Pyrimidines are abundant in the meteorite, but they are quite unlike the pyrimidines of living systems.

The Oparin-Haldane hypothesis of chemical evolution postulates that the ubiquitous biological nitrogenous bases such as uracil, thymine, cytosine, adenine and guanine will be formed preferentially by the flux of energy through a primitive atmosphere. Why we could not find biological heterocyclic compounds in the Murchison meteorite is a mystery. It should be noted, however, that neither the biological nitrogenous bases nor the bases found in the Murchison meteorite have ever been detected in simple spark-discharge experiments on methane, ammonia and water. It appears that the Murchison pyrimidines were most likely synthesized extraterrestrially by an abiotic process, but that this process did not directly generate the "biological" bases. Three other carbonaceous chondrites, the Murray, Orgueil and Allende meteorites, have been found to contain heterocyclic molecules similar to the ones in the Murchison meteorite.

Hydrocarbons held the center of the stage in earlier studies of organic compounds in meteorites. Hydrocarbons are inherently more stable than molecules incorporating nitrogen and oxygen. As a result hydrocarbons are more likely to remain intact during the rigors of cosmic time. By 1963 it was known that meteorites contain hydrocarbons with linear structures (aliphatic hydrocarbons) and ring structures (aromatic hydrocarbons). Over the next six years Oró also demonstrated the presence of linear isoprenoid hydrocarbons, a class of molecules with biological significance. For example, isoprenoid hydrocarbons can be derived from chlorophyll. Oró showed that at least nine carbonaceous chondrites contain small concentrations of such molecules. Although it was hoped that these isoprenoid hydrocarbons might have been synthesized by extraterrestrial chemical processes, the bulk of the evidence supported the conclusion that they originated with contamination by terrestrial biological substances.

Since samples of the Murchison meteorite appeared to be free of terrestrial amino acids and pyrimidines, it seemed possible that the hydrocarbons in the meteorite might also be free of terrestrial contaminants. The gas chromatogram of the aliphatic hydrocarbons indicated that the meteorite contained a complex mixture of molecules that could not be separated into individual peaks. Mass spectrometry of the mixture revealed that the dominant compounds were complex molecules with a double-ring structure. No linear isoprenoid hydrocarbons were found.

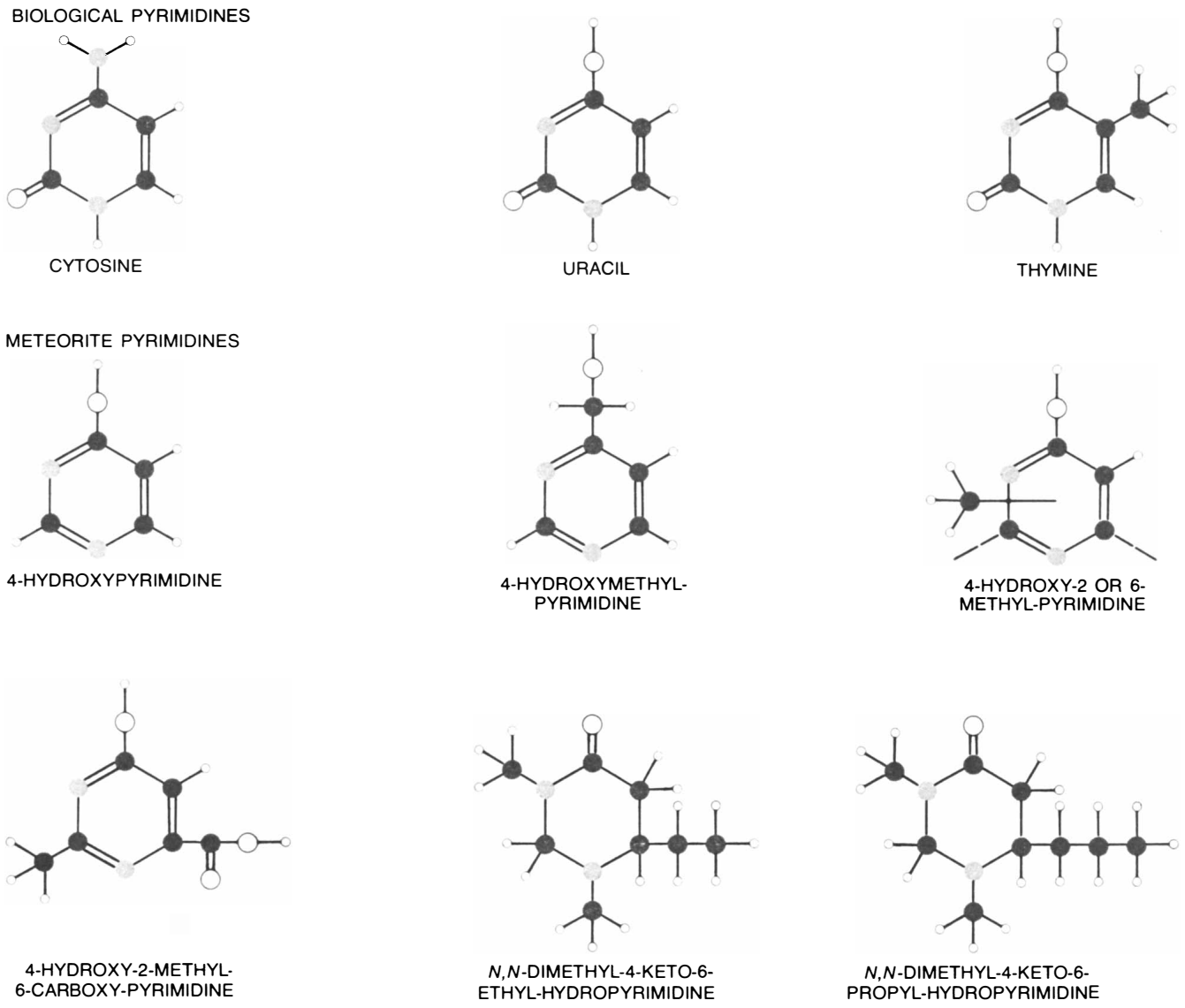
Experiments testing the hypotheses of chemical evolution have produced a series of hydrocarbons bearing a remarkable resemblance to the series found in the Murchison meteorite. Aliphatic hydrocarbons were synthesized when methane (CH_4) was subjected to an electric discharge. The gas-chromatographic pattern of these hydrocarbons looked very much like the pattern produced by the hydrocarbons from the meteorite. More important, in the synthetic mixture

of hydrocarbons the complex molecules with a double-ring structure dominated the mixture just as they did in the meteorite.

Although aromatic hydrocarbons had been reported in a number of carbonaceous chondrites, the Murchison meteorite provided an additional opportunity to look at these compounds. Again gas chromatography and mass spectrometry were called on, and they revealed that at least 23 aromatic hydrocarbons are present. The mixture of aromatic hydrocarbons in the meteorite is much simpler than the mixture of aliphatic compounds. The aromatics are similar in structure to those produced in chemical-evolution experiments where methane was reacted at a very high temperature.

One additional kind of measurement has been performed on the Murchison meteorite. Various fractions of carbon compounds were analyzed for their composition of the various isotopes of carbon by Ian Kaplan and his group at the University of California at Los Angeles. The ratio of the isotopes carbon 13 to carbon 12 was analyzed and compared with the ratio obtained from a widely used terrestrial standard: carbonate from the Pee Dee belemnite, a fossil of a type found in South Carolina. When measured against this standard, organic extracts of terrestrial sediments and petroleum have an isotopic ratio of carbon 13 to carbon 12 between 25 and 35 parts per 1,000 less than that of the standard. Organic extracts from the Murchison meteorites show ratios between 4.4 and 5.9 parts per 1,000 greater than the ratio of the standard. Clearly the extracts from the meteorite are in a range distinct from the organic substances associated with geological materials. The carbonate carbon in the meteorite has an isotopic ratio of 45.5 parts per 1,000 greater than the standard, which compares well with the ratios obtained for carbonates in other carbonaceous chondrites.

All the determinations made so far with the Murchison meteorite indicate that its organic compounds come from extraterrestrial sources, and that they are most likely of abiotic origin. Thus the recent results obtained from the Murchison chondrite and other carbonaceous chondrites provide a new basis for chemical evolution and a further impetus for the search for extraterrestrial life. Future studies of Mars and Jupiter should provide additional information about organic compounds present in other portions of our solar system, and should lead to further increases in our understanding of organic cosmochemistry.



PYRIMIDINES were found in four carbonaceous chondrites, but these compounds (*bottom two rows*) differed significantly in structure from pyrimidines commonly occurring in biological materials (*top row*). In simple electric-spark discharge experiments designed to test hypotheses of chemical evolution neither the meteoritic pyrimidines nor biological pyrimidines have been detected.

We want to be useful ...and even interesting

Doing anything July 10th?

Those for whom a total solar eclipse makes a day very important professionally have quite definite plans already for July 10, 1972 and quite definite opinions on what's worth doing that day. The path of totality being what it is through the oft-clouded skies of Alaska, the Northwest Territories, Quebec, and the Maritimes, mere enthusiasts may find themselves luckier. We are told that an opportunity presents itself to confirm photographically the existence of comets or, conceivably, other small celestial objects much nearer the sun than Mercury. We are also told that it will take lots of luck indeed, and lots of coolness, patience, and focal length.

If interested and if there is still time, send to Dept. 916, Kodak, Rochester, N.Y. 14650 for "How to Enjoy the Eclipse."

Spontaneity through information density

We hope that you (or at least your family) have learned in recent weeks of Kodak's pocket principle in photography: film only 16 mm wide in a one-inch-thick camera leads to high-grade 3½″x4½″ snapshots, 5″x7″ enlargements, or a fine slide show of unusual spontaneity in expression and behavior.

This principle rests on a bewildering interlace of more basic principles in engineering, chemistry, economics, and business law. Most basic of all is the limit on information density set by chemical technology. To suggest that the limit has now been even approached would be self-serving pessimism. A physicist among us gives guidance on what improvements would be noticeable.

Author of a doctoral thesis on fourth sound in HeII, he finds that translating spatial frequency concepts into migration behavior of chemical entities in a photographic color emulsion adequately utilizes his professional capabilities.

The photographic chemists he counsels think like chemists. To deal with the photographic image in terms of Fourier analysis is the way of a communications engineer. The physicist breaks down a photograph into several million data points and processes the data to enhance some spatial frequencies.

See, for example, how an electron micrograph of a cell nucleus can be thus processed to accentuate the spatial frequencies of the mitochondria:

BEFORE

AFTER

Takes too much computer time for routine use. Feasible when a study that correlates subjective judgments of sharpness with patterns of spatial frequency enhancement can be supported by millions of camera-buying spontaneity-seekers.

Micropublishing takes great leap forward, attaché cases to lighten

For delivery soon after October 1, 1972, we now offer a $95 microfiche reader. To book order, find Kodak in the Yellow Pages under "Microfilming."

It weighs less than five pounds. Its 12V lamp—autodome type—needs no fan. Soundless. Smart optical design, for a bright, contrasty front-projected image. A minor purchase for the home or office, costing less than a comfortable chair to put in front of it.

If publishers and learned societies can sell the idea that ink is not the vital ingredient in disseminating specialized information and according recognition to authors, prospects for economic viability may brighten. Now it should be easier to sell that idea.

Price subject to change without notice.

The KODAK EKTALITE 120 Reader can be used like this:

Other than scholarly pursuits will be served. This guy could just as well be reading computer-compiled reports from his home office, a year's worth of which on paper would cost $95 just to mail to him. For three bucks more than that price, this reader plugs into the cigar lighter in a salesman's car or an equipment service truck. Think of the volume of information that could be waiting in the parking lot during a call!

Kodak

Economic viability is desirable

IF YOU'RE STILL AIRLINE BE THIS IS WHAT

First we introduced Ambassador Service* on our long flights across the United States.

To give you more reasons to fly with us than with any other airline.

Then we expanded Ambassador Service to our international flights to Europe and across the Pacific.

And now, we've introduced a whole new service on many of our short business flights. TWA's new Ambassador Express.* It's on over 250 flights a day in 16 cities across the United States.

No other airline does so many of these things in so many places as TWA.

AMBASSADOR SERVICE ACROSS THE UNITED STATES, TO EUROPE AND THE PACIFIC.

AMBASSADOR EXPRESS BETWEEN MAJOR MIDWEST AND EAST COAST CITIES.

*Service mark owned exclusively by TWA.

THE NEW L-1011. TWA is adding to its fleet the quietest, most modern plane in the world. The new L-1011. It will be in service this summer.

747 COACH LOUNGE. On all our 747 flights across the United States.

707 COACH LOUNGE. On nearly every 707 Ambassador flight across the U.S.A. No other airline has it.

REFURNISHED PLANES. On all 707 Ambassador Service flights you'll find completely refurnished interiors. New seats, carpets, fabrics, colors, etc. We're also redoing the interiors of our planes used for Ambassador Express flights, and by mid-summer they'll all be completed.

FLYING SOME SIDES TWA, YOU'RE MISSING.

CHOICE OF MEALS. On Ambassador Service meal flights you'll find a choice of 3 international meals in coach, 5 in first class. On Ambassador Express meal flights, you can choose between a full meal or a light snack.

CHOICE OF MOVIES.[1] You'll find a choice of 2 first-run movies on every movie flight across the U.S.A. (Nominal charge for movies in coach; movies are shown on flights departing after 11 AM and before 10 PM.)

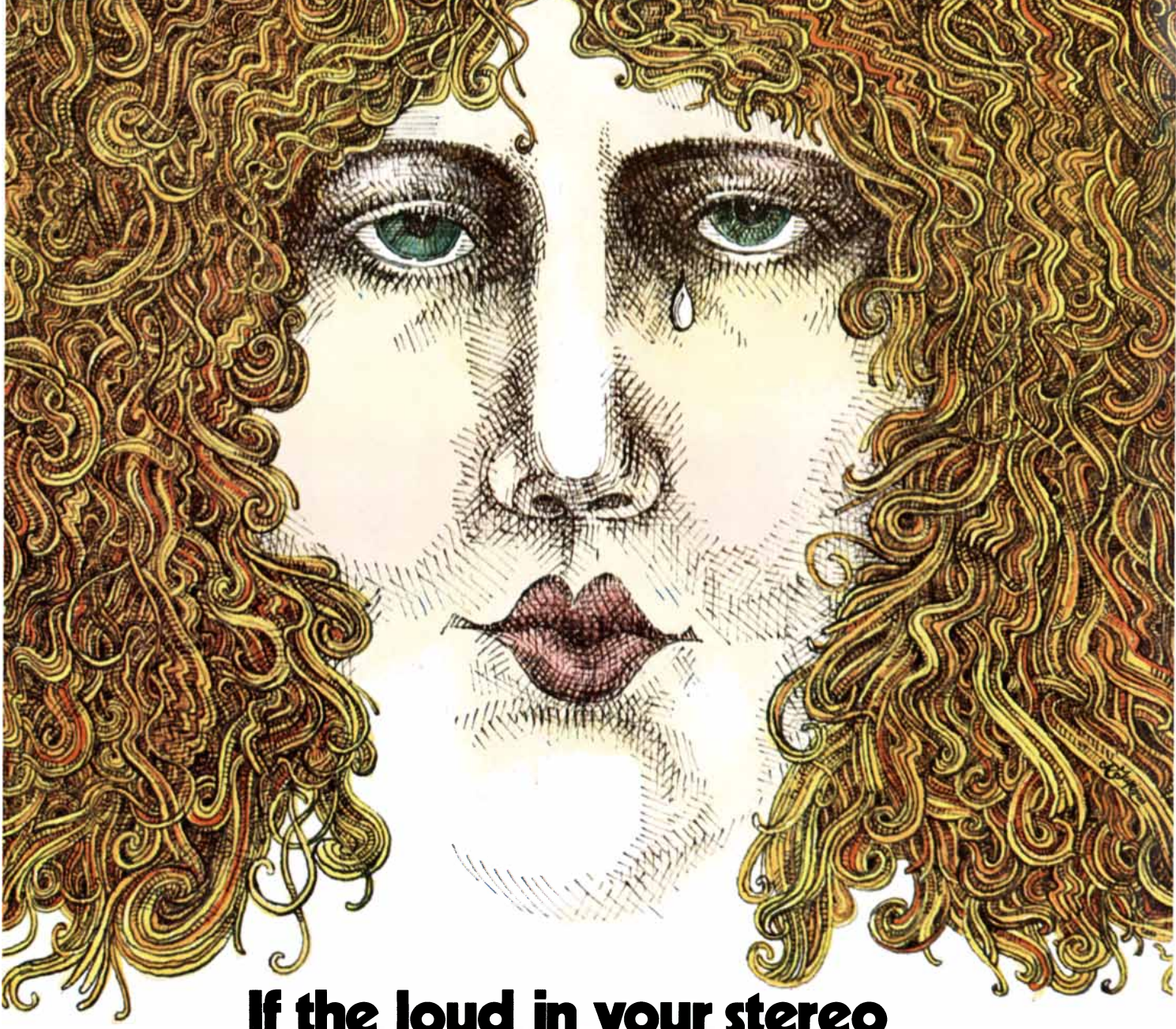
NEW LUGGAGE COMPARTMENTS. In the coming months, we'll be putting new luggage compartments on nearly every one of our Ambassador Express flights, and other short business flights. So you can bring even a full-size suitcase right on board with you.

TWIN SEATS. On 707 Ambassador Service flights you'll find the Twin Seat in coach, and now we're also putting them in our 727s for short business flights. If the plane's not full, our Twin Seat can be 3 across, 2 across or even a couch.

CONTAINERIZED BAGGAGE. On Ambassador Service flights across the United States, or to Europe, we guarantee you closed containerized baggage.

TWA's AMBASSADOR SERVICE AND AMBASSADOR EXPRESS. THE MOST COMFORTABLE FLEET OF PLANES IN THE WORLD.

1. Movies by Inflight Motion Pictures, Inc.



If the loud in your stereo gives her a pain in the head get a Marantz.

It's not that she's got super-duper sensitive hearing (like that spotted beagle two doors down), it's because most women hear better than men, so when she screams turn down the sound what she really means is turn down the damn distortion because the distortion is driving her bananas.

Not so with Marantz stereo. Take the Marantz 2270 stereo AM/FM receiver for $549.95. It delivers a walloping 140 watts RMS power at less than 0.3% distortion...which means virtually NO DISTORTION. And because Marantz measures distortion at continuous full power through the whole listening range—it won't bring tears to her eyes or a pain to her head. NO MATTER HOW LOUD. Other companies measure power and distortion only in middle listening range. So they quote their highest power and lowest distortion only in the middle. But on the low and high sides their power is down and their distortion is up. So if you pay for 140 watts be sure you get 140 watts at both ends and in the middle.

With Marantz you get exactly what you pay for. If all you need is 30 watts, take our model 2215 for $249.95. Want twice as much power? Our model 2230 gives you 60 watts RMS for $349.95. And our model 2245 at $449.95 delivers 90 watts RMS. ALL FULL RANGE CONTINUOUS POWER.

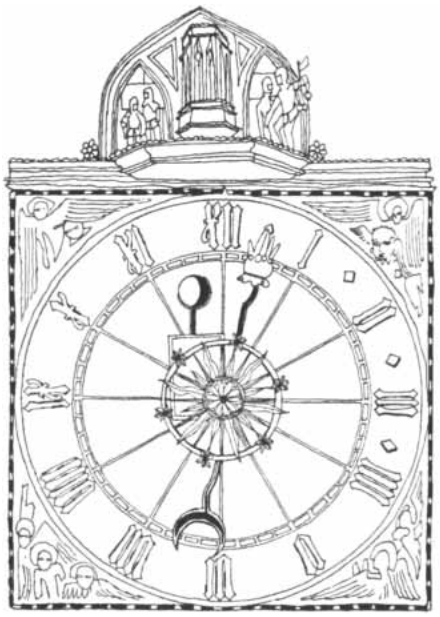
No matter which model you choose, remember this. You're getting Marantz quality. The same quality that goes into the magnificent Model 19 FM stereo receiver (shown) that costs $1200. YES. That's right, a cool $1200! But it is the absolute, ultimate, very best there is.

Visit your Marantz dealer. And take your wife along. (Or that spotted beagle two doors down.)

marantz®
We sound better.

©Marantz Co., Inc., a subsidiary of Superscope, Inc., P.O. Box 99M, Sun Valley, Calif. 91352. In Europe: Marantz Int'l. S.A., Brussels, Belgium. In Canada: Electrohome, Ltd., Kitchener, Ontario. Prices subject to change without notice. Send for free catalog.

SCIENCE AND THE CITIZEN



The Mysterious Moon (Cont.)

Rock samples, magnetometer readings and seismic signals provided by *Apollo 16* intensify rather than clarify puzzles about the structure of the moon. It had been thought that the ninth and 10th lunar explorers, John W. Young and Charles M. Duke, Jr., would find the landing site near the crater Descartes (lunar longitude nine degrees south, latitude 15 degrees 31 minutes east) strewn with volcanic rocks largely undisturbed since they had crystallized billions of years ago. The site was selected because it represented a typical highland region in which craters appear to be flooded with lava-like material. Instead nearly all the rocks returned to Houston are breccias, rocks consisting of many small particles welded by heat and pressure into a conglomerate. Unlike the breccias found at other sites, however, most of the *Apollo 16* breccias have a very white matrix speckled with very dark inclusions that seem to have a simpler chemistry than those seen in other samples. Selenologists who favor a volcanic origin for the highland rocks are now proposing that the *Apollo 16* samples are "igneous breccias" not unlike those sometimes produced by volcanoes on the earth. An opposing school sees no reason to believe that the *Apollo 16* breccias are any different from those found at other sites: breccias formed by the impact of large meteorites. Thus the role played by volcanism in the moon's early history remains obscure.

Using a portable magnetometer, Young and Duke found regions where the magnetic field at the lunar surface was some three times higher than the highest value measured at other sites, which had already seemed difficult to explain. Although the new magnetic values (about 350 gammas) are far below those at the earth's surface, the lunar magnetic field is produced entirely by the remanent magnetism in the rocks, whereas the earth's field is provided chiefly by an active "dynamo" in the earth's core. The clear implication is that when the lunar rocks solidified about three billion years ago, they were immersed in a magnetic field of at least 1,000 gammas, a field one-fiftieth as strong as the earth's present field.

There seem to be only three possible sources for a magnetic field of that intensity, and there are substantial objections to each. The moon may have had its own internal dynamo, but this implies a molten core that could hardly have cooled to the temperature now believed to exist in the center of the moon. (One hypothesis is that a barely molten core of iron sulfide could just satisfy the temperature restrictions.) Alternatively, the sun may have generated an exceptionally strong field at the time the lunar rocks were solidifying, but that field, unlike the sun's present field, would have had to be unidirectional. Finally, the moon may have been so close to the earth (within three or four earth radii) that the cooling lunar rocks were magnetized by the earth's field. The moon may indeed have been close enough to the earth at one period, but no theory of earth-moon dynamics provides for the moon's remaining close enough for the time needed to produce the observed magnetization, a time measured in hundreds of millions of years.

A seismic experiment designed to test a new model of the moon's crustal structure ran into unexpected difficulty. The intention was to have the S-IVB stage of the launch vehicle crash into the moon midway among three seismometers installed at other Apollo sites. Unfortunately contact with the S-IVB was lost, leaving the precise time and location of impact uncertain. Nevertheless, from the seismic signals recorded at the three stations it was possible to deduce the landing site and impact time. Since the site was evidently 1,050 kilometers from the *Apollo 15* instrument, however, the signal there was the weakest yet recorded from an artificial impact, leaving the seismologists in doubt whether or not it actually represented the initial compression, or *P*, wave. Assuming that it was the initial wave, it traveled about 10 percent slower than had been predicted by the model. The model had assumed a sound velocity of nine kilometers per second in the lunar mantle, a velocity indicated by earlier seismic readings at the *Apollo 12* and *Apollo 14* sites. If this velocity were shown to be typical for large areas of the moon, seismologists would be hard pressed to characterize the composition of the material in the lunar mantle. The actual sound velocity inferred from the flawed experiment was about eight kilometers per second, or about the same value observed in the upper part of the earth's mantle. The original intention was to crash the S-IVB stage of *Apollo 17* on the back of the moon; now, in all likelihood, that idea will be abandoned so that the *Apollo 16* experiment can be repeated.

The first use of the moon as a serious astronomical observatory was highly successful. A special *f*/1 Schmidt camera built at the Naval Research Laboratory, capable of recording ultraviolet radiation between 800 and 1,500 angstroms, produced 209 photographs of the earth's geocorona, or hydrogen halo, and of selected stellar and galactic objects, chiefly in the southern hemisphere of the sky as viewed from the earth. The pictures of the earth confirm that it is surrounded out to at least 15 earth radii by a halo of hydrogen, which under bombardment by sunlight emits Lyman-alpha radiation at a wavelength of 1,216 angstroms.

Core Encore

In the summer of 1968 an unusual research vessel sailed from Texas on the first of a series of voyages that have now covered more than 100,000 nautical miles. Equipped to take core samples of marine sedimentary deposits lying many hundreds of feet below the sea floor by means of drill strings that are sometimes more than four miles long, the *Glomar Challenger* in the past four years has

brought to the surface for study a total of 60,000 feet of cores from more than 300 holes drilled in the bottom of the Atlantic, Pacific and Indian oceans and the Mediterranean, Caribbean, Arabian and Red seas. The deep-sea drilling project, which has helped to revolutionize fundamental concepts of geology, was scheduled to end this August. The National Science Foundation, the source of project funds, has now announced that the *Glomar Challenger* will continue drilling until at least 1975.

Information gleaned from the sea-floor cores includes concrete evidence of the geological process of crustal renewal and destruction known as sea-floor spreading. The ocean basins of today evidently came into existence no earlier than some 160 million years ago; the most ancient bottom rock found by the *Glomar Challenger* is of this relatively recent age, in contrast to continental rocks that are well over three billion years old. The spreading process that formed the ocean basins is documented in other cores; these samples indicate that the age of the sea floor increases proportionately with distance from the mid-ocean ridges where crustal upwelling takes place. Both findings add significantly to the existing array of geophysical evidence for sea-floor spreading and continental drift (see "Plate Tectonics," by John F. Dewey; SCIENTIFIC AMERICAN, May).

The drilling has produced evidence of vertical earth movements as well as horizontal ones. For example, a major submarine ridge that runs for 1,500 miles along the floor of the Indian Ocean now lies at a depth of more than a mile below the surface. Cores containing peat and oyster shells, however, show that the ridge was once at the surface, forming a chain of islands complete with marshes and lagoons.

Cores taken in the Mediterranean contained what may be the most surprising *Glomar Challenger* discovery of all. It appears that no more than six or eight million years ago, in the Pliocene epoch, most of this historic sea was transformed from an arm of the world ocean into a waterless waste where the process of evaporation formed thick deposits of salt. A million years or so later the Mediterranean basin filled again; it has contained a body of water with depths up to two miles ever since.

The *Glomar Challenger* is owned and operated by a Los Angeles firm, Global Marine, Inc., that engages in petroleum exploration. The ship's drillers must have had a sense of nostalgia when in 1968, taking a core from one of the Sigsbee Knolls, a series of domelike underwater formations at a depth of two miles in the Gulf of Mexico, they struck oil. The crew sealed the drill-hole with concrete to prevent oil pollution and sailed on.

Vision Chips

A little more than two years ago two engineers at the Bell Telephone Laboratories, W. S. Boyle and G. E. Smith, announced that they had discovered a new approach to the handling of information with semiconductor devices. Basically their idea, which they named charge-coupling, was to store information in the form of electric charge inside a semiconducting material and to move it around by means of voltages applied to electrodes at the surface of the semiconductor. In principle their invention opened the way to an entire new family of electronic devices capable of performing certain complex information-handling tasks better and more efficiently than such tasks could be performed with existing semiconductor devices.

The realization of the charge-coupling concept in the form of practical applications has come with surprising speed. Within the past few months Bell Laboratories engineers have turned out a variety of new devices incorporating the charge-coupling principle. Several of the experimental devices were described recently by Smith at the spring meeting of the American Physical Society in Washington.

The charge-coupling approach appears to be particularly promising for image sensors of various types. For example, a new type of solid-state television camera based on the use of a charge-coupled image sensor has been built at Bell Laboratories. It produces a television-like image without the cumbersome vacuum tube, electron beam and high voltage required by conventional television systems.

The charge-coupled image sensor consists of a flat chip of oxidized silicon with a complex array of metal electrodes. Accumulations of electric charge are built up in the silicon by the absorption of light from an image focused on the chip. The amount of charge accumulated in each area is proportional to the amount of incident light, making the charge pattern an electric analogue of the image. By varying the voltages on the electrodes one can "freeze" the charge pattern in the silicon and control its movement, making it possible to shift the charge packets one by one to an output electrode. There they embody an analogue video signal corresponding to the light variations of the original image. Further development of this experimental charge-coupled video system is expected to lead to a television camera that is smaller, simpler, more sensitive, more reliable and less expensive than existing cameras.

Among the other applications of the charge-coupling principle currently under study by Smith and his colleagues is a high-density computer memory capable of storing 13,675 units of analogue or digital information on a single silicon chip. This device, called a shift register, incorporates essentially the same integrated-circuit chip as the image sensor at the heart of the solid-state television camera does. A closely related charge-coupled device built recently at Bell Laboratories is a 500-storage-element analogue delay line capable of delaying television signals for more than 100 microseconds. This device's potential uses include echo suppression in transmission links, bandwidth reduction in television transmission, filter network synthesis and other aspects of electronic signal-processing. Another application in this area is a 500-element linear scanning device that can be used for high-resolution imaging of graphic materials such as type and photographs.

Computer Talk

One of the most widely discussed long-range prospects of computer technology has been that one day large computers over a large area would be interconnected so that they could communicate with one another. Such a system is now in existence, and it is expanding rapidly. It has been set up by the Advanced Research Projects Agency of the Department of Defense for users involved in research financed by the Government. By last month the network, which is called ARPANET, included 24 computer centers ("nodes" in the terminology of the system), ranging from Massachusetts to California and having from one to four digital computers each for a total of 31.

Since the computers in the network were made by several different manufacturers and have different modes of operation, the key problem in establishing the network was to make the inputs and outputs of the machines compatible. The problem was solved with an interface message processor (IMP in ARPANET terminology) at each node. An IMP is a relatively small computer that accepts messages from as many as five host computers at the site, breaks each message into 1,000-bit packets and

sends each packet toward its destination (another IMP or a terminal interface processor, which is for groups that do not have their own computer) over the communication line that has the least traffic at the time. The system has leased a number of communication lines from various commercial carriers; each line has a capacity of 50,000 bits per second. At the receiving end the IMP processes incoming messages so that they can be received by the computer to which they are addressed. The system has a response speed of .1 second for a question and .3 second for a one-page answer.

ARPANET has been set up under the supervision of Lawrence G. Roberts, director for information-processing techniques of the Advanced Research Projects Agency. Looking to a time when commercial networks similar to ARPANET will be available, he thinks "many institutions will find it far more economic" to be part of such a system than to operate their own computer center. Hardware-sharing will be the motive for establishing a commercial network, according to Roberts, but thereafter he sees a progression to "data-base sharing—direct retrieval from remote, one-of-a-kind data bases" in three or four years; software sharing (the use of programs not available on the subscriber's primary computer) in four to eight years, and text-oriented services (communicating with libraries or handling office paperwork) in 10 to 20 years.

The Case of the Missing Neutrinos

It is generally believed that the sun's radiant energy originates with thermonuclear reactions deep in the solar interior in which hydrogen nuclei are fused into the nuclei of heavier atoms. One product of these reactions should be a flood of neutrinos: massless uncharged particles that interact so little with other particles that the outer layers of the sun and solid bodies such as the earth are virtually transparent to them. Raymond Davis, Jr., of the Brookhaven National Laboratory has devised an experiment to detect solar neutrinos, measure their flux and their energy and thus test the theory of the solar interior. Davis' first results in 1968 showed that the flux of neutrinos was between four and eight times lower than expected. Present theory of the solar interior predicts that neutrinos are produced at such a rate that Davis' detector should capture two neutrinos per day. Even after improvements in the sensitivity of the apparatus, results of the experiment over the past two years show that the capture rate is less than .2 neutrino per day, a tenth the number predicted.

Davis' detector, buried a mile under solid rock in the Homestake Gold Mine in Lead, S.D., is a huge tank containing 100,000 gallons of the common dry-cleaning solvent tetrachloroethylene. The solvent is 85 percent chlorine; when a neutrino is absorbed by an atom of chlorine, an atom of the radioactive isotope argon 37 is formed. At intervals of about 100 days the contents of the tank are swept with helium gas to remove the argon 37 present, which is then detected by its radioactive decay in a proportional counter.

Explanations of the discrepancy between Davis' results and the theory of the solar interior are varied, but none is very satisfactory. It is possible that the internal structure or the internal composition of the sun is not well understood. Perhaps the energy-generating process is cyclic and the cycle is now at a minimum. Or perhaps some neutrinos decay before they reach the earth. The problem is certain to receive intense study in the near future.

The Male Presence

The age at menarche, or first menstrual period, for girls in the U.S. and in Europe has declined at the rate of three to four months per decade for the past 100 years. Although the trend toward earlier sexual maturity is amply documented, little is known about the causal factors. Investigators have tried to relate the onset of menstruation to climate, altitude, season, nutrition, hygiene, disease, race and social and psychological factors. Because many of the factors are interrelated, however, it has not been possible to determine their individual effects. Now J. G. Vandenbergh and his associates at the North Carolina Department of Mental Health report an interesting animal model of accelerated sexual maturation that may help to identify some of the dominant factors.

In earlier studies Vandenbergh had found that when young female mice that had not yet attained puberty were housed with an adult male mouse, they first ovulated 20 days earlier than young female mice housed without a male present. The apparent cause was stimulation by a pheromone: a substance secreted by one member of a species that affects the behavior of other members of the same species. When soiled bedding from cages housing males was added to cages housing isolated young females, the sexual maturation of the females was accelerated. On the other hand, some investigators have found that sexual maturation in mice was dependent on the amount of protein in the diet or on reaching a critical body weight. Vandenbergh and his colleagues decided to conduct an experiment to determine the relative importance of these factors.

Female mice were weaned at 21 days of age and placed in individual cages. They were exposed to three levels of male stimulation: (1) an adult male mouse lived with a young female, (2) soiled bedding from adult-male cages was added daily to the cage of a young female and (3) young females were maintained with no male contact. Three levels of dietary protein were also tested: 8 percent, 16 percent and 24 percent. Each diet contained equal amounts of fats, vitamins and minerals.

The findings, presented in *Journal of Reproductive Fertility,* show a relation between the level of dietary protein and the age of sexual maturation. Female mice fed a diet with 8 percent protein matured later than female mice reared on higher levels of protein. Moreover, of the females housed with males only 69 percent of the mice reared on 8 percent protein produced litters, whereas 81 percent of those on 16 percent protein and 100 percent of those on 24 percent protein bore young.

The presence of a male, however, advanced the age of first ovulation on the average to about eight days after weaning. Females exposed only to male bedding began to ovulate about 13 days after weaning; isolated females began about 15 days after weaning. Analysis of variance indicated that although both dietary protein and the presence of males (or their odor) were significant factors in regulating sexual maturation, the male presence contributed 47.3 percent of the total variance and diet accounted for only 4.8 percent. Body weight appeared to have no influence on the onset of ovulation.

Vandenbergh and his colleagues also found that female mice housed in groups with a male ovulated later than female mice housed singly with males. Although the effect of male stimulation is strong under both conditions, the presence of other females appears to inhibit sexual maturation. Moreover, female mice housed singly and not exposed to a male matured sexually at an earlier age than females in a group exposed to a male did. Vandenbergh and his associates observe: "[The fact that] social stimulation was more effective than protein intake may have relevance to the phenomenon of accelerated sexual maturation in human females."

THE EVOLUTION OF REEFS

The community of plants and animals that builds tropical reefs is descended from an ecosystem of two billion years ago. The changes in this community reflect major events in the history of the earth

by Norman D. Newell

To a mariner a reef is a hazard to navigation. To a skin diver it is a richly populated underwater maze. To a naturalist a reef is a living thing, a complex association of plants and animals that build and maintain their own special environment and are themselves responsible for the massive accumulation of limestone that gives the reef its body. The principal plants of the reef community are lime-secreting algae of many kinds, including some whose stony growths can easily be mistaken for corals. The chief animal reef-builders today are the corals, but many other marine invertebrates are important members of the reef community.

This association of plants and animals in the tropical waters of the world is the most complex of all ocean ecological systems; as we shall see, it is also the oldest ecosystem in earth history. Its closest terrestrial counterpart, in terms of organization and diversity, is the tropical rain forest. Both settings evoke an image of exceptional fertility and exuberant biomass. Both are dependent on light in much the same way; the sunlight filters down through a stratified canopy, and the associations at each successive level consist of organisms whose needs match the available illumination and prevailing conditions of shelter. There is even a parallel between the birds of the rain forest and the crabs and fishes of the reef. Both play the part of lords and tenants, yet their true role in the history and destiny of the community is essentially passive.

It is a common belief that a reef consists principally of a rigid framework composed of the cemented skeletons of corals and algae. In reality more than nine-tenths of a typical reef consists of fine, sandy detritus, stabilized by the plants and animals cemented or otherwise anchored to its surface. Physical and biochemical processes that are little understood quickly convert this stabilized detritus into limestone. The remains of dead reef organisms make a substantial contribution to the detritus. This major component of the reef, however, has a fabric quite different from the upward-growing lattice of stony algal deposits and intertwined coral skeletons that forms the reef core.

The Reef Community

The interaction of growth and erosion gives the reef an open and cavernous fabric that in an ecological sense is almost infinitely stratified and subdivided. In the dimly lit bottom waters at the reef margin, rarely more than 200 feet below the surface, caves and overhanging ledges provide shelter for the plants and animals that thrive at low levels of illumination. From the bottom to the surface is a succession of reef-borers, cavern dwellers, predators and detritus-feeders, each living at its preferred or obligatory depth, that includes representatives of nearly every animal phylum. Near and at the surface the sunlit, oxygen-rich and turbulent waters provide an environment that contributes to a high rate of calcium metabolism among the myriads of reef-builders active there.

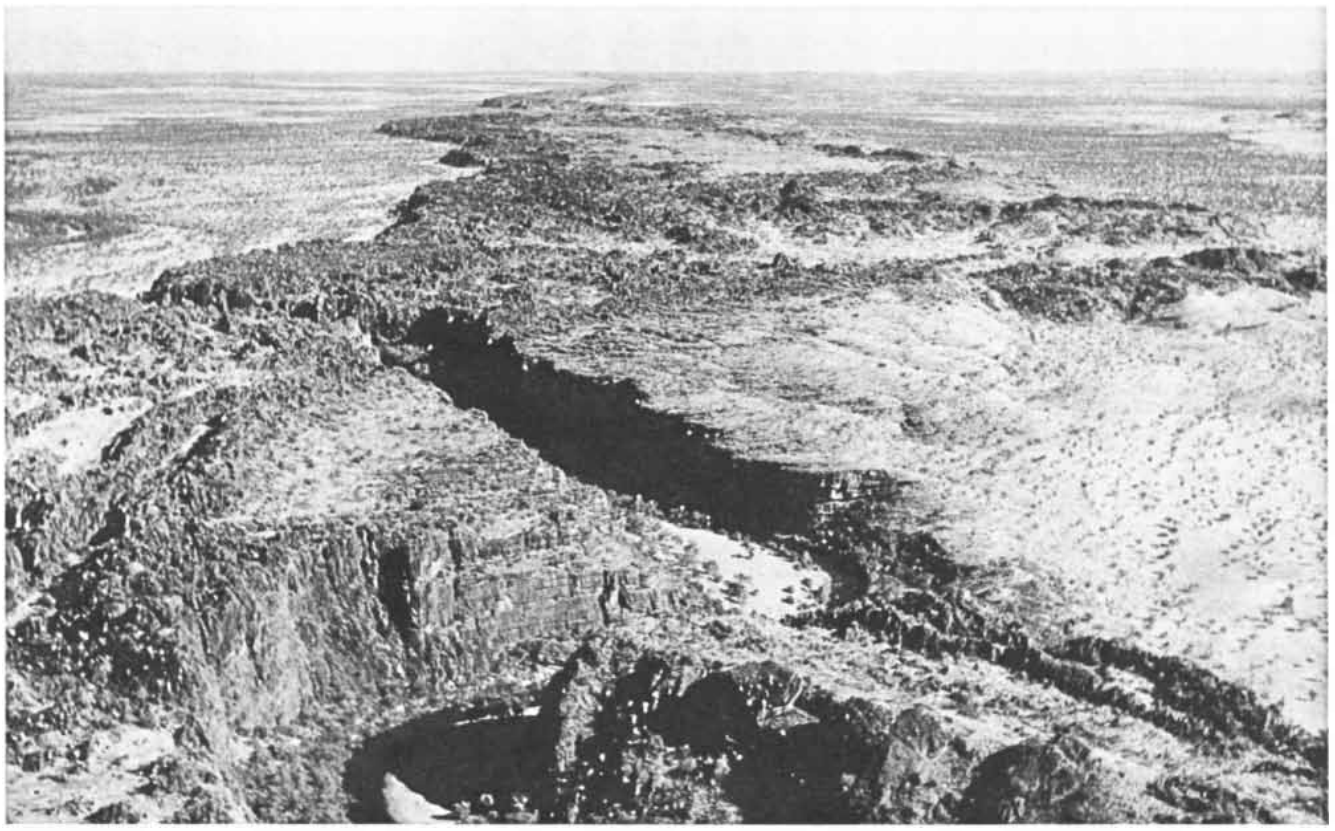
The most familiar of the reef animals, the corals, are minute polyps that belong to the phylum Coelenterata. The polyps live in symbiosis with zooxanthellae, microscopic one-celled plants embedded in the animals' tissue, where they are nourished by nitrogenous animal wastes and, through photosynthesis, add oxygen to the surrounding water. Experiments show that the zooxanthellae promote the calcium metabolism of the corals. The corals themselves are carnivores; they feed mainly on small crustaceans and the larvae of other reef animals.

The limestone-secreting algae—blue-green, green and red—are the principal food base of the reef community, just as the plant life ashore nourishes terrestrial herbivores. The algae are distributed across the reef in both horizontal and vertical zones. The blue-green algae are most common in the shallows of the tidal flat, an area where red algae are absent. The green algae are predominantly back-reef organisms and the reds are mostly reef and fore-reef inhabitants [*see illustration on page 64*].

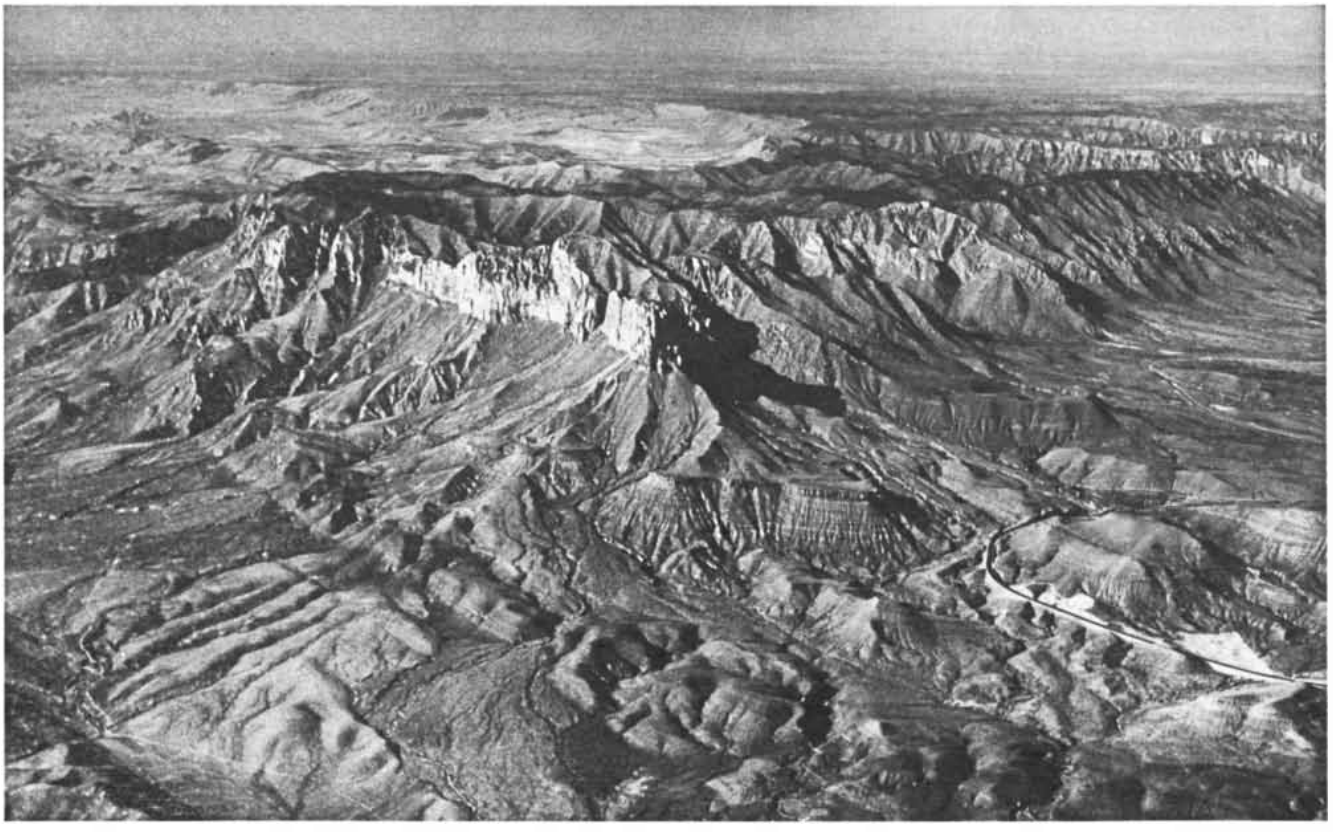
The other important members of the reef community are all animals. Next in significance to the corals as reef-builders are several limestone-secreting families of sponges, members of the phylum Porifera. The phylum Protozoa is represented by a host of foraminifera species whose small limy skeletons add to the deposits in and around reefs. Several species of microscopic colonial animals of the phylum Bryozoa also contribute their limestone secretions, as do the spiny sea urchins and elegant sea lilies of the phylum Echinodermata, the bivalves of the phylum Brachiopoda and such representatives of the phylum Mollusca as clams and oysters, all of whose accumulated skeletons and shells contribute to the reef limestones.

Many organisms in the community do not add significantly to the reef structure; some burrowers and borers are even destructive. The marine worms that inhabit the reef are soft-bodied and thus incapable of contributing to the reef mass. The hard parts of such reef dwellers as crabs and fishes are systematically consumed by scavengers. A few fragments may escape, but except for such passive and minor contributions to the reef detritus these organisms are not reef-builders.

The reef community is adapted to a low-stress environment characterized



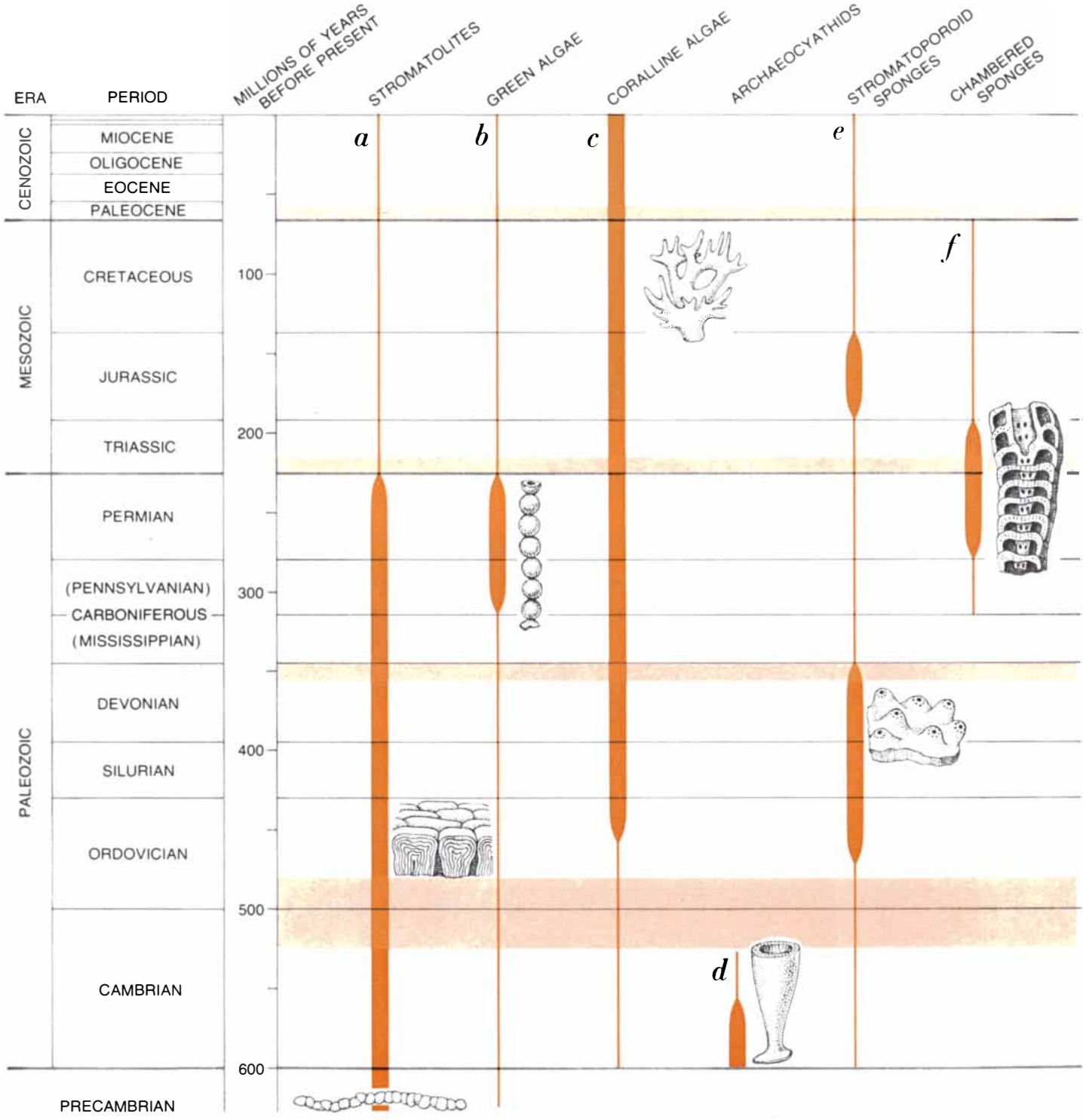
FOSSIL BARRIER REEF was built in upper Devonian times by a community of marine plants and animals living in the warm seas that covered part of Australia more than 350 million years ago. Exposed by later uplift and erosion, the reef now forms a belt of jagged highlands, known as the Napier Range, in Western Australia. A stream has cut a canyon (*foreground*) through the reef rock.



YOUNGER FOSSIL REEF, built some 250 million years ago in the Permian period, forms a rim of rock 400 miles long surrounding the Delaware Basin (*right*) on the border between Texas and New Mexico. Most of the reef is buried under later deposits, but one exposure forms this 40-mile stretch of the Guadalupe Mountains. El Capitan (*foreground*), a part of the reef front, is 4,000 feet high.

by the absence of significant seasonal change. The mean winter temperature of the water where reefs grow is between 27 and 29 degrees Celsius and the difference between summer and winter monthly mean temperatures is three degrees C. or less. The water is clear (so that the penetration of light is at a maximum), agitated (so that it is rich in oxygen) and of normal salinity. Even under these ideal circumstances many reef organisms (for example corals) do not grow at depths greater than 65 feet. This adaptation to freedom from stress makes the community remarkably sensitive to environmental change.

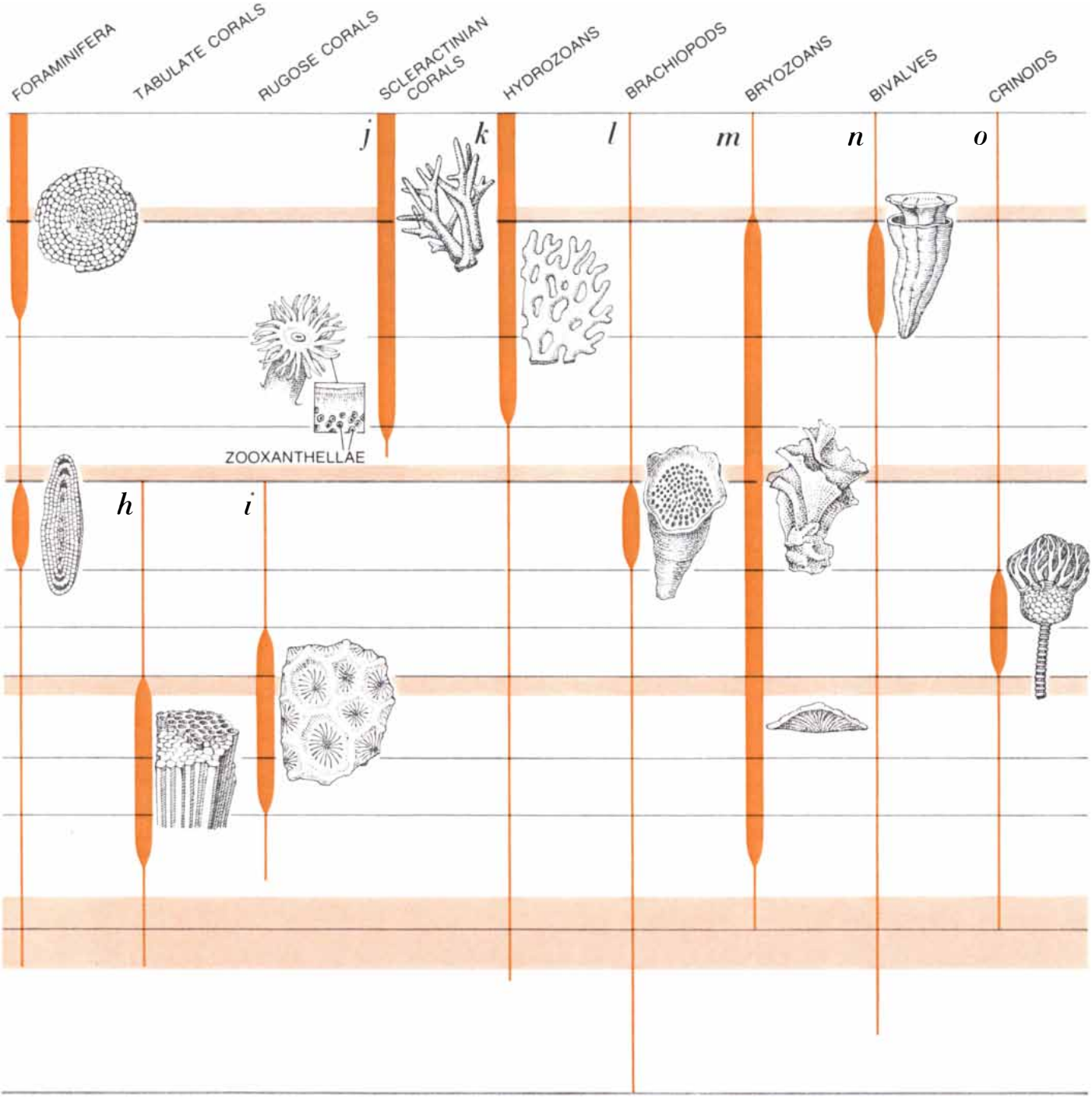
The fossil record documents hundreds of episodes of sweeping mass extinction, some continent-wide and others worldwide. These times of ecological disruption have simultaneously affected such disparate organisms as ammonites at sea and dinosaurs ashore, the plants on land and the protozoans afloat among the



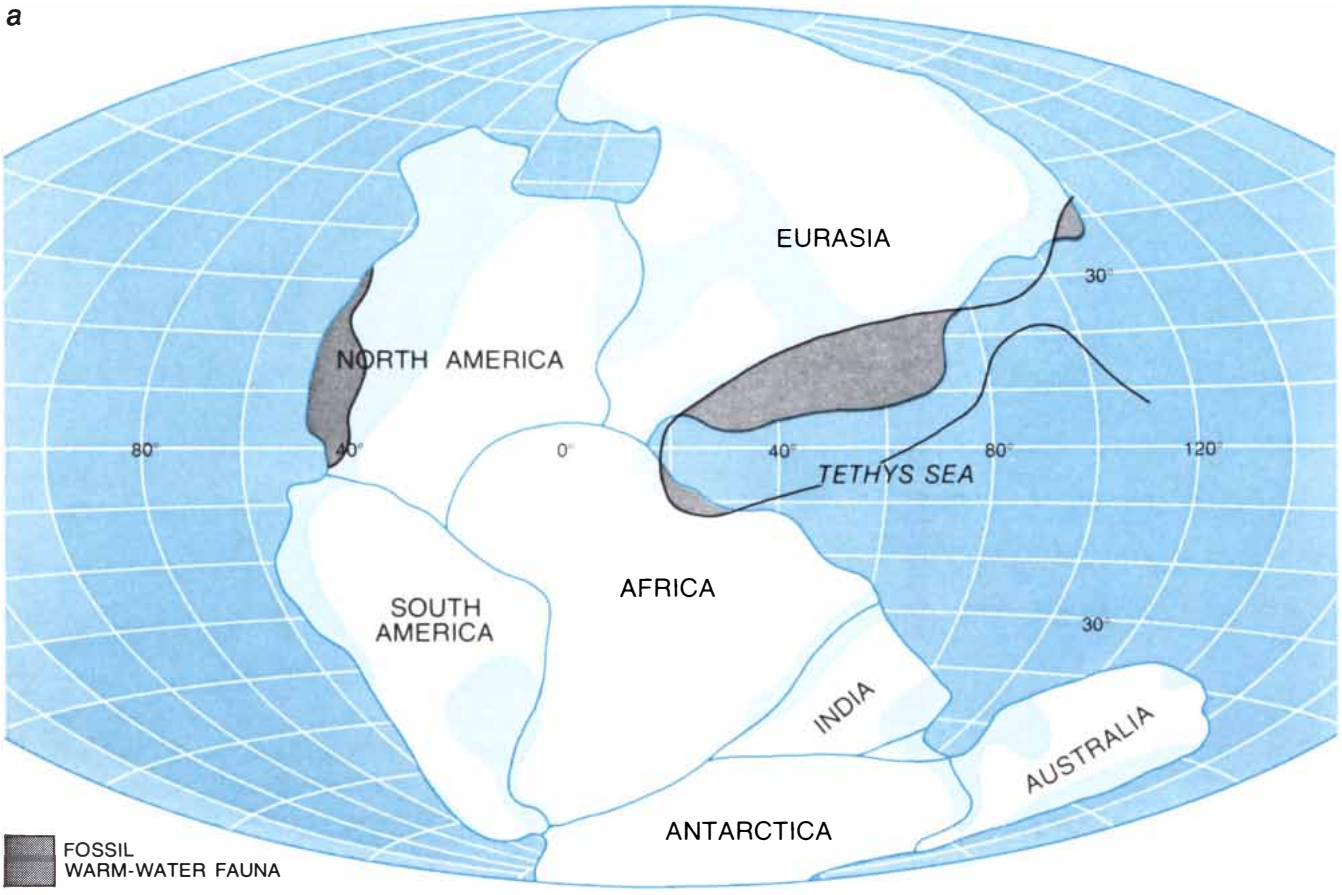
FOUR COLLAPSES (*bands of color*) **have altered the composition of the reef community since the initial association between plant and animal reef-builders was established nearly 600 million years ago. This was when a group of spongelike animals, the archaeocyathids** (*d*), **appeared among the very much older reef-forming algal stromatolites** (*a*) **at the start of the Paleozoic era. In less than 70 million years the archaeocyathids became extinct; their demise marks the first community collapse. A successor community arose in mid-Ordovician times. Its members included coralline algae** (*c*); **the first corals, tabulate** (*h*) **and rugose** (*i*); **stromatoporoid sponges** (*e*), **and communal bryozoans** (*m*). **The group flourished almost to the end of the Devonian period, some 350 million years ago, the**

oceanic plankton. The causative phenomena underlying the disruptions must therefore be unlike the ordinary, Darwinian causes of extinction—natural selection and unequal competition—that tend to affect species individually and not en masse.

For generations, in a reaction to the biblical doctrine of catastrophism that dominated 18th-century geology, scholars have viewed the apparent lack of continuity in fossil successions with skepticism. The breaks in the record, they proposed, were attributable to inadequate collections or to accidents of fossil preservation. At the same time, certain pioneers—T. C. Chamberlin and A. W. Grabau in the U.S. and Hans Stille in Germany—saw the breaks in fossil continuity as reflections of real events and sought a logical explanation for them. These men were eloquent proponents of a theory of rhythmic pulsations within the earth: diastrophic move-



time of the second collapse. Its successor, some 13 million years later, contained a new sponge group (*f*) and increased numbers of green algae (*b*), foraminifera (*g*), brachiopods (*l*) and crinoids (*o*). These reef-builders flourished until the end of the Paleozoic era and the third collapse. The next resurgence occupied most of the Mesozoic era. It was marked by the appearance of modern corals (*j*) and a dramatic upsurge of a mollusk group, the rudists (*n*), that became extinct at the time of the fourth community collapse some 65 million years ago. Draining of shallow seas during the Cenozoic era produced cooler climates and led to formation of the Antarctic ice cap. Both developments have been factors in restricting the successor community in diversity and distribution.



TWO FACTORS that have affected world geography and climate are the movement of the continental plates and their greater or lesser invasion by shallow seas. The status of both factors during three key intervals in earth history is shown schematically here and on the opposite page. Near the end of the Paleozoic era (*a*) the continental plates had gathered into a single land area. Many of the reef-building species were then pantropical in distribution. By the end of the Mesozoic era (*b*) sea-floor spreading had separated the continental plates. The Atlantic Ocean had become enough of a barrier to the migration of reef organisms between the Old and New World to allow the evolution of species unique to each region. At the end of the Mesozoic era the shallow seas encroaching on the continents were drained completely. Early in the Cenozoic era (*c*) the shallow seas had been reestablished in certain zones, but the total continental area above water was larger than in the Mesozoic. The change resulted in a trend toward greater seasonal extremes of climate; the distribution of tropical and subtropical organisms, however, remained quite broad, as palm-tree fossils show.

ments that had been accompanied by significant fluctuations in sea level and consequent disruptive changes in climate and environment.

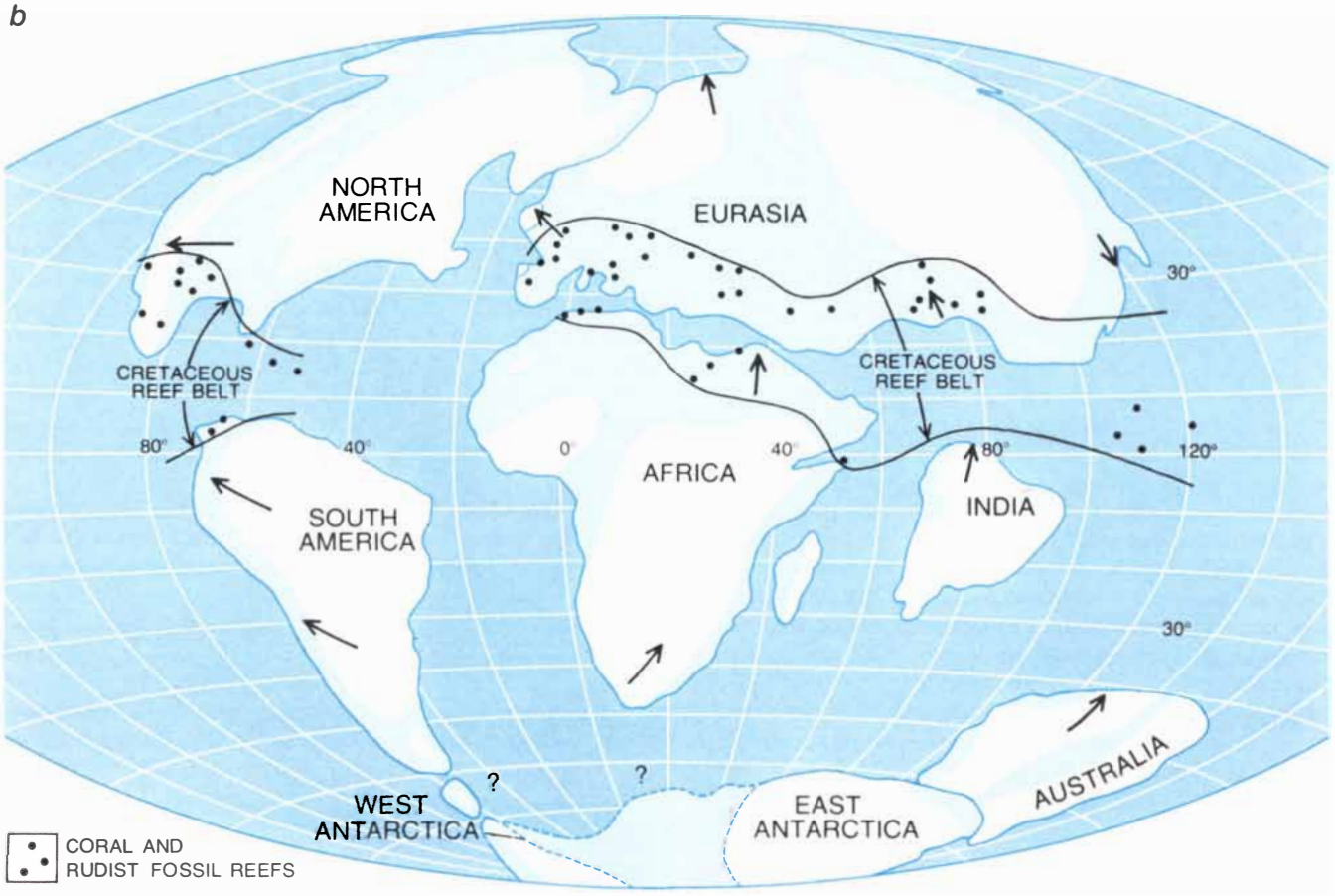
The lack of any demonstrable physical mechanism that might have produced such simultaneous worldwide geological revolutions kept a majority of geologists and paleontologists from accepting the proposed theory of the origins of environmental cycles. Today, however, we have in plate tectonics a demonstrated mechanism for changes in sea level and shifts in the relative extent of land and sea such as Chamberlin and his colleagues were unable to provide [see "Plate Tectonics," by John F. Dewey; SCIENTIFIC AMERICAN, May]. Significant changes in the volume of water contained in the major ocean basins, produced by such plate-tectonic phenomena as alterations in the rate of lava welling up along the deep-ocean ridges or in the rate of sea-floor spreading, have resulted sometimes in the emergence and sometimes in the flooding of vast continental areas. The changing proportions of land and sea meant that global weather patterns alternated between mild maritime climates and harsh continental ones.

Now, reef-building first began in the earth's tropical seas at least two billion years ago. As we have seen, the modern reef community is so narrowly adapted to its environment as to be very sensitive to change. It seems only logical to expect that the same was true in the past, and that changes in reef communities of earlier times would faithfully reflect the various rearrangements of the earth's land masses and ocean basins that students of plate tectonics are now documenting. The expectation is justified; whereas many details of earth history will be clarified only by future geological and paleontological research, the record of fossil reef communities accurately delineates a number of the main catastrophic episodes.

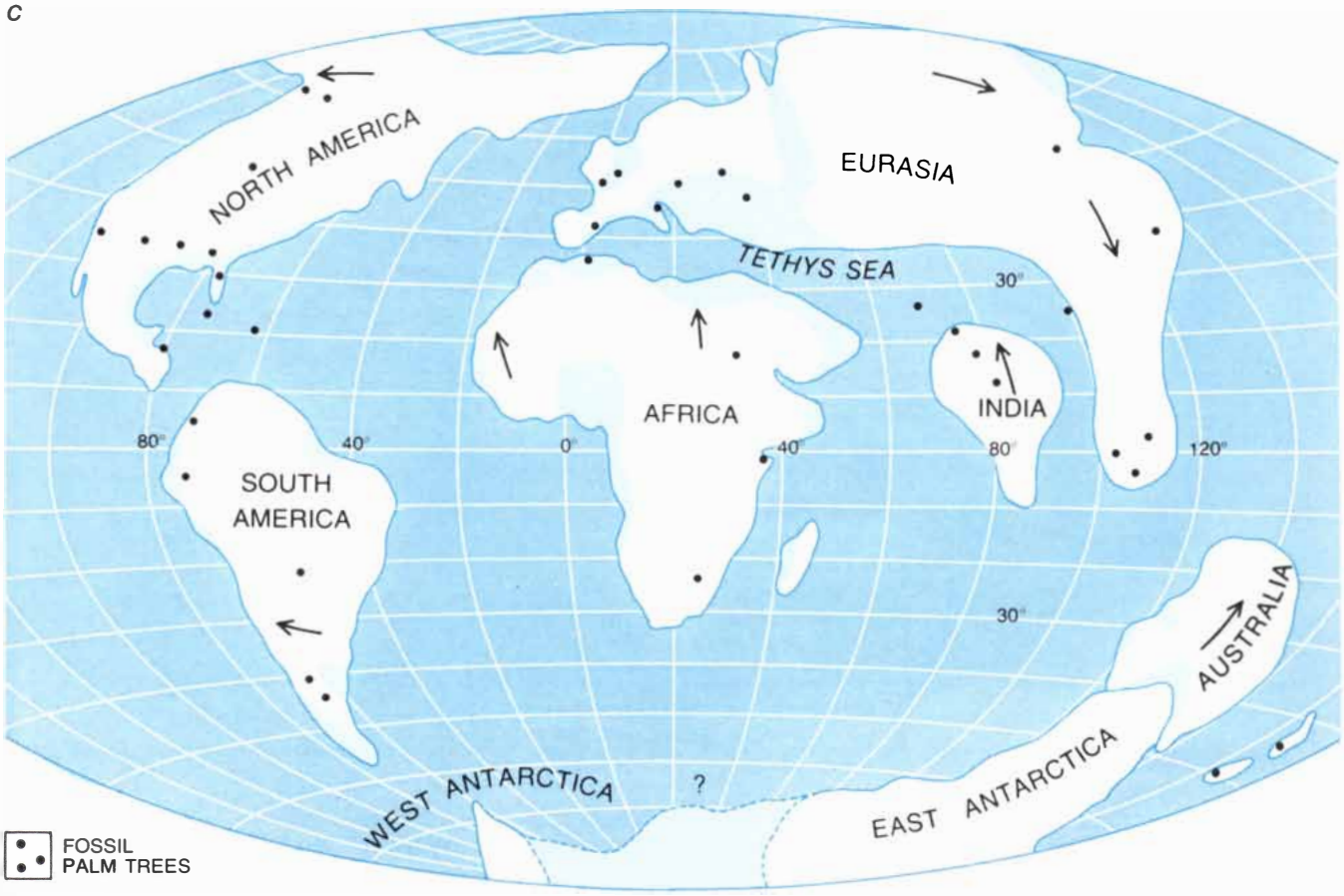
As one might expect, the oldest of all types of reef is the simplest. Algae alone, without associated animals, are responsible for limestone reef deposits, billions of years old, that were formed in the seas of middle and late Precambrian times.

The Precambrian algae produced extensive accumulations of a distinctively laminated limestone that are found, flanked by aprons of reef debris, in rock formations around the world. Geologists call these characteristic limestone masses stromatolites, from the Greek for "flat" and "stone." The microscopic organisms that built the stromatolites are rarely

b



c



preserved as fossils, but they must surely have been similar to the filamentous blue-green algae that form similar masses of limestone today.

The accomplishments of these Precambrian algal reef-builders were not inconsiderable: individual colonies grew upward for tens of feet. They did so by trapping detrital grains of calcium carbonate and perhaps by precipitating some of the lime themselves. The resulting fossil bodies take the form of trunklike columns or hemispherical mounds.

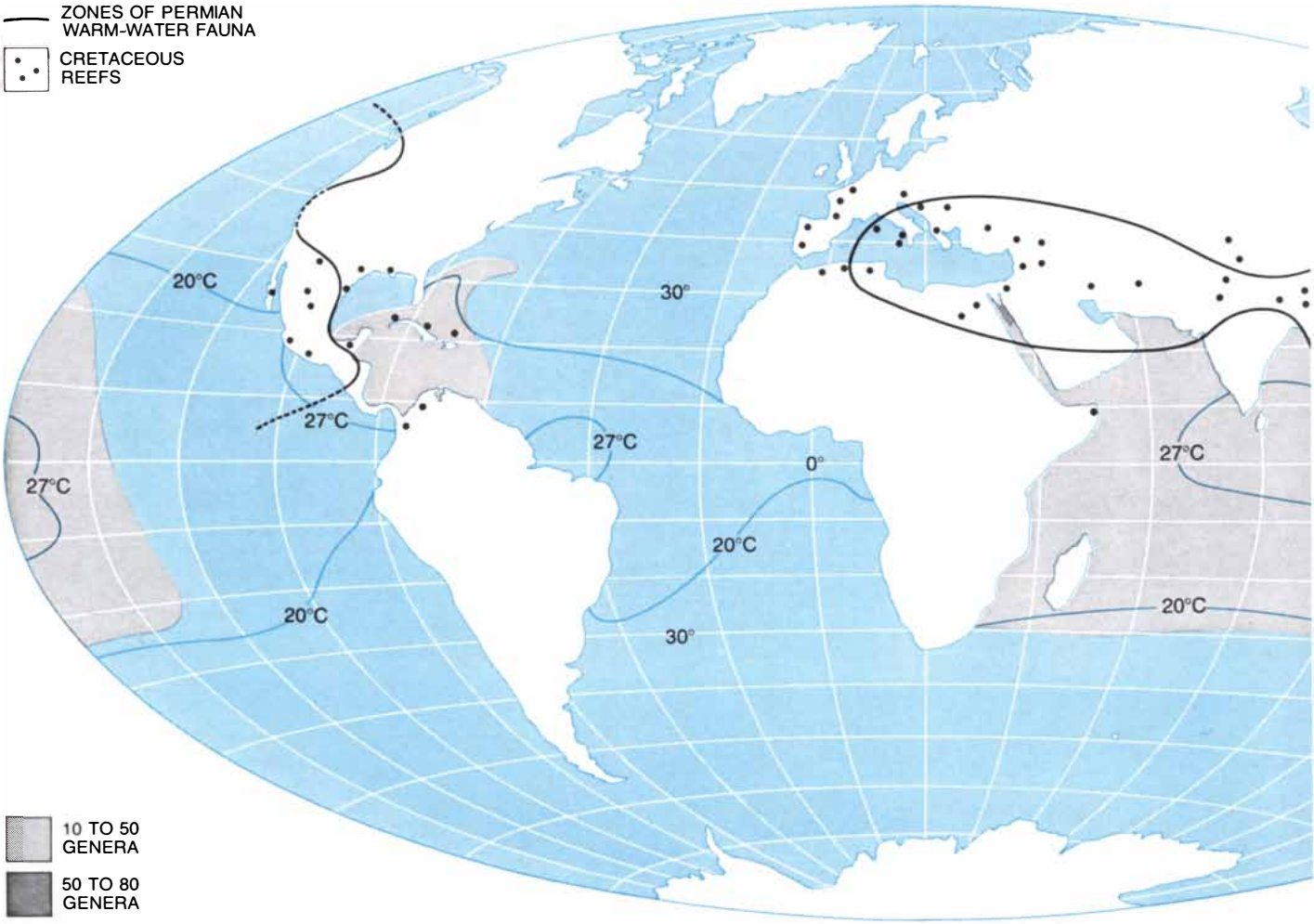
As I outline the evolution of the reef community the reader will note that I often speak of first and last appearances. This does not mean, of course, that the organism being discussed was either instantly created or instantly destroyed. Each had an extensive evolutionary heritage behind it when some chance circumstance provided it with an appropriate ecological niche. Similarly, the decline and extinction of many major groups of organisms can be traced over periods of millions of years, although in numerous instances the time involved was too short to be measured by the methods now available.

Enter the Animals

The long Precambrian interval ended some 600 million years ago. The opening period of the Paleozoic era, the Cambrian, saw the first establishment of a reef community. The stromatolites' first partners were a diverse group of stony, spongelike animals named archaeocyathids (from the Greek for "ancient" and "cup"). In early Cambrian times these stony animals rooted themselves along the stromatolite reefs, grouped together in low thickets or scattered like shrubs in a meadow. It is not hard to imagine that the vacant spaces within and between these colonies provided shelter for the numerous bottom-feeding trilobites that inhabited the Cambrian seas. Not every reef harbored archaeocyathids, however; some reefs of early and middle Cambrian times are composed only of stromatolites.

By the end of the middle Cambrian, some 540 million years ago, the archaeocyathids had vanished. No single cause for this extinction, the first of the four major disasters to overtake the reef community, can be identified. One imaginable cause—competition from another reef-building animal—can be ruled out completely. The seas remained empty of any kind of reef-building animal throughout the balance of the Cambrian and until the middle of the Ordovician period, some 60 million years later. All reefs built during this long interval were the work of blue-green algae alone.

Fossil formations in the Lake Champlain area of New York, a region that lay under tropical seas in middle Ordovician times some 480 million years ago, contain the first evidence of a renewed association between reef-building plants and animals. The community that now arose was a rather complex one. Stromatolites continued to flourish and a second kind



REEF-BUILDERS TODAY are confined to a narrow belt, mainly between 30 degrees north and 30 degrees south latitude (*light gray*), and even within this belt the greatest number of species are found where the minimum average water temperature is 27 degrees Celsius. Paleozoic and Mesozoic reef fossils, however, are found in areas far outside today's limits (*black lines and dots*). This suggests that the true Equator in those times lay well to the north of the equators shown on the maps on the preceding pages. The asym-

of plant life appeared: the coralline red alga *Solenopora*, a direct progenitor of the modern coralline algae. Colonial bryozoans, previously insignificant in the fossil record, now assumed an important role in the expanding reef community. Animal newcomers included a group of stony sponges, the stromatoporoids, some shaped like encrusting plates and others hemispherical or shrublike in form. These calcareous sponges were to play a major role in the community for millions of years. The most significant new animals, however, in light of subsequent developments, were certain stony coelenterates: the first of the corals. The intimate collaboration between algae and corals, apparently unknown before middle Ordovician times, has continued (albeit with notable fluctuations) to the present day.

These new arrivals and the other corals that appeared during the Paleozoic era were mainly of two types. In one type the successive stages of each polyp's upward growth were recorded as a series of parallel floors that subdivide the stony tube sheltering the animal; these organisms are called tabulate corals. The conical or cylindrical stone tube that sheltered the second type has conspicuous external growth wrinkles on its surface; these corals are called rugose.

A little more than 350 million years ago, near the end of the Devonian period, worldwide environmental changes caused a number of mass extinctions. Among the victims were many previously prominent marine organisms, including several groups in the reef community. That community now underwent a major retrenchment. Up to this point the tripartite association between algae, sponges and corals that first appeared in Ordovician times had proliferated for 130 million years without significant disturbance. The environmental alterations that nearly wiped out the reef community at the end of that long and successful period of radiation remain unidentified, although one can conjecture that a change from a mild maritime climate to a harsh continental one probably played a part. In any event the episode was severe enough so that only scarce and greatly impoverished reef communities, consisting in the main of algal stromatolites, survived during the next 13 million years. Not until well after the beginning of the Carboniferous period was there a community resurgence.

Some 115 million years passed between the revival of the reef community in Carboniferous times and the end of the Paleozoic era; the interval includes most of the Mississippian and all of the Pennsylvanian (the two subdivisions of the Carboniferous) and the closing period of the Paleozoic, the Permian. The revitalized assemblage that radiated in the tropical seas during this time continued to include stromatolites, numerous bryozoans and brachiopods and a dwindling number of rugose corals. Except for these organisms, however, the community bore no great resemblance to its predecessor of the middle Paleozoic. Both the stromatoporoid sponges and the tabulate corals are either absent from Carboniferous and Permian reef deposits or are present only in insignificant numbers.

Two new groups of calcareous green algae, the dasycladaceans and codiaceans, now attained quantitative importance in the reef assemblage. As if to match the decline of the stromatoporoid sponges, a second poriferan group—the calcareous, chambered sphinctozoan sponges—entered the fossil record. At the same time a group of echinoderms—the crinoids, or sea lilies—assumed a larger role in the reef community. As the Paleozoic era drew to a close, the crinoids and the brachiopods achieved their greatest diversity; their skeletons preserved in Permian reef formations number in the thousands of species.

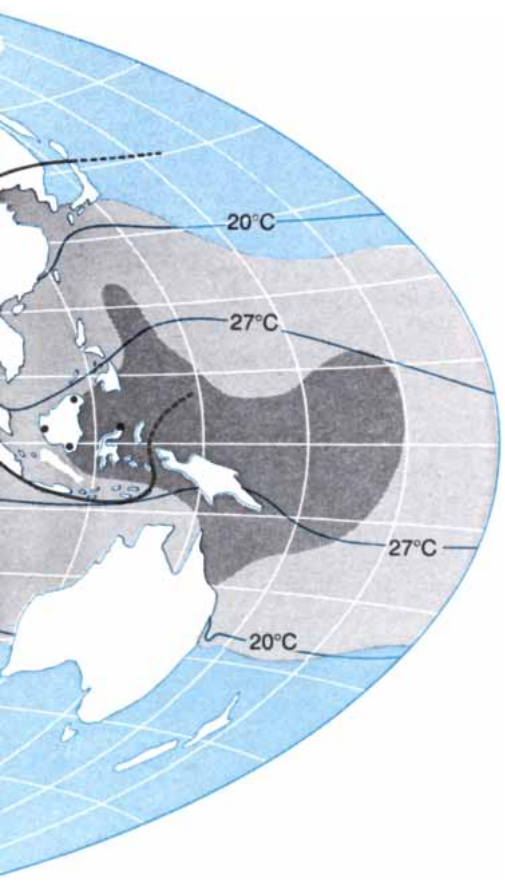
The Third Collapse

Half of the known taxonomic families of animals, both terrestrial and marine, and a large number of terrestrial plants suffered extinction at the end of the Paleozoic era. The alteration in environment that occurred then, some 225 million years ago, had consequences far severer than those of Devonian times. In the reef community the second successful radiation—based principally on a new tripartite association involving algae, bryozoans and sphinctozoan sponges—came to an end; reefs are unknown anywhere in the world for the first 10 million years of the Mesozoic era.

What was the cause of this vast debacle? There is little enough concrete information, but analogy with later and better-understood events encourages the conjecture that once again unfavorable changes in climate and habitat were major factors. In late Paleozoic times all the continents had come together to form a single vast land mass: Pangaea. Continental ice sheets appeared in the southern part of this supercontinent, a region known as Gondwana, in Carboniferous and early Permian times. These glaciers are concrete evidence of cooler climates; paleomagnetic studies show that the glaciated areas were then all located near the South Pole. Any relation between these early Permian ice sheets and the widespread late-Permian extinctions, however, is not yet evident. What is probably more significant is evidence that, at least during a brief interval, all the shallow seas that had invaded continental areas were completely drained at the end of the Paleozoic era. Serious climatic consequences must have resulted from the disappearance of a mild, primarily maritime environment.

In late Paleozoic times a wide tropical seaway, the Tethys, almost circled the globe. The only barrier to the Tethys Sea was formed by the combined land masses of North America and western Europe, which were then connected. The tongue of the Tethys that eventually became the western Mediterranean constituted one end of the seaway. The opposite end invaded the west coast of North America so deeply that great Permian reefs arose in what is now Texas.

The Mediterranean extremity of the Tethys Sea was the setting for a signifi-



metry of the fossil-reef belt with respect to the present belt suggests that much of a once wider array has been engulfed by subduction along continental-plate boundaries.

cant development when, after 10 million years of eclipse, the reef community was once again revitalized. There, in mid-Triassic times, a new group of corals, the scleractinians, made its appearance. The scleractinians were the progenitors of the more than 20 families of corals living in the reef community today. At first the new coral families, six in all, were represented in only a few scattered reef patches found today in Germany, in the southern Alps and in Corsica and Sicily. Even by late Triassic times, some 200 million years ago, the new corals were still subordinate as reef-builders to the calcareous algae.

The Mesozoic Community

During the 130 million years or so of Jurassic and Cretaceous times the reef community once again thrived in many parts of the world. The stromatoporoid sponges, all but extinct since the community collapse of Devonian times, returned to a position of some importance during the Jurassic. The new coral families steadily increased in diversity and reached an all-time peak in the waters bordering Mediterranean Europe during the Cretaceous period. In that one region there flourished approximately 100 genera of scleractinians; this is a greater number than can be found worldwide today. The reef community in this extremity of the Tethys Sea was also rich in other reef organisms. It included the two groups of sponges and such reef-builders as sea urchins, foraminifera and various mollusks. In addition a hitherto minor group of coralline red algae, the lithothamnions, now began to play an increasingly important role. By this time the stromatolites, important reef-builders throughout the Paleozoic, were no longer conspicuous members of the reef community.

Early in the Cretaceous period, some 135 million years ago, there was an unfavorable interval: reefs are unknown in the fossil record for some 20 million years thereafter. This pause merely set the stage, however, for a further efflorescence. Both in the Mediterranean area and in the waters of the tropical New World, hitherto unknown or insignificant coral families appeared. The present Atlantic Ocean was just beginning to form. Regional differences between reef communities in the Old World and the New that appeared at this time testify to the growing effectiveness of the Atlantic deeps as a barrier to the ready migration of reef organisms.

An extraordinary evolutionary performance was now played. Certain previously obscure molluscan members of the reef community, bivalves known as rudists, abruptly came into prominence as primary reef-builders. The next 60 million years saw a phenomenal rudist radiation that brought these bivalves to the point of challenging the corals as the dominant reef animals. Along the sheltered landward margins of many fringing barrier reefs rudists largely supplanted the corals. Their cylindrical and conical shells were cemented into tightly packed aggregates that physically resembled corals, and many of the aggregates grew upward in imitation of the coral growth pattern. Before the end of the Cretaceous period some 65 million years ago the rudists had attained major status in the reef community. Then at the close of the Cretaceous they quite abruptly died out everywhere.

Ever since the pioneer days of geology in the 18th century the end of the Cretaceous period has been known as a great period of extinctions. Nearly a third of all the families of animals known in late Cretaceous times were no longer alive at the beginning of the Cenozoic era. The reef community was not exempt; in addition to the rudists, two-thirds of the known genera of corals died out at this time. Forms of marine life other than reef organisms were also hard hit. The ammonites, long a major molluscan group, had suffered a decline during the final 10 million years of the Cretaceous; by the end of the period they were all extinguished. The belemnites, another major group of mollusks, declined sharply, as did the inoceramids, a diverse and abundant group of clams that had previously flourished worldwide. Among the foraminifera, the free-floating, planktonic groups suffered in particular.

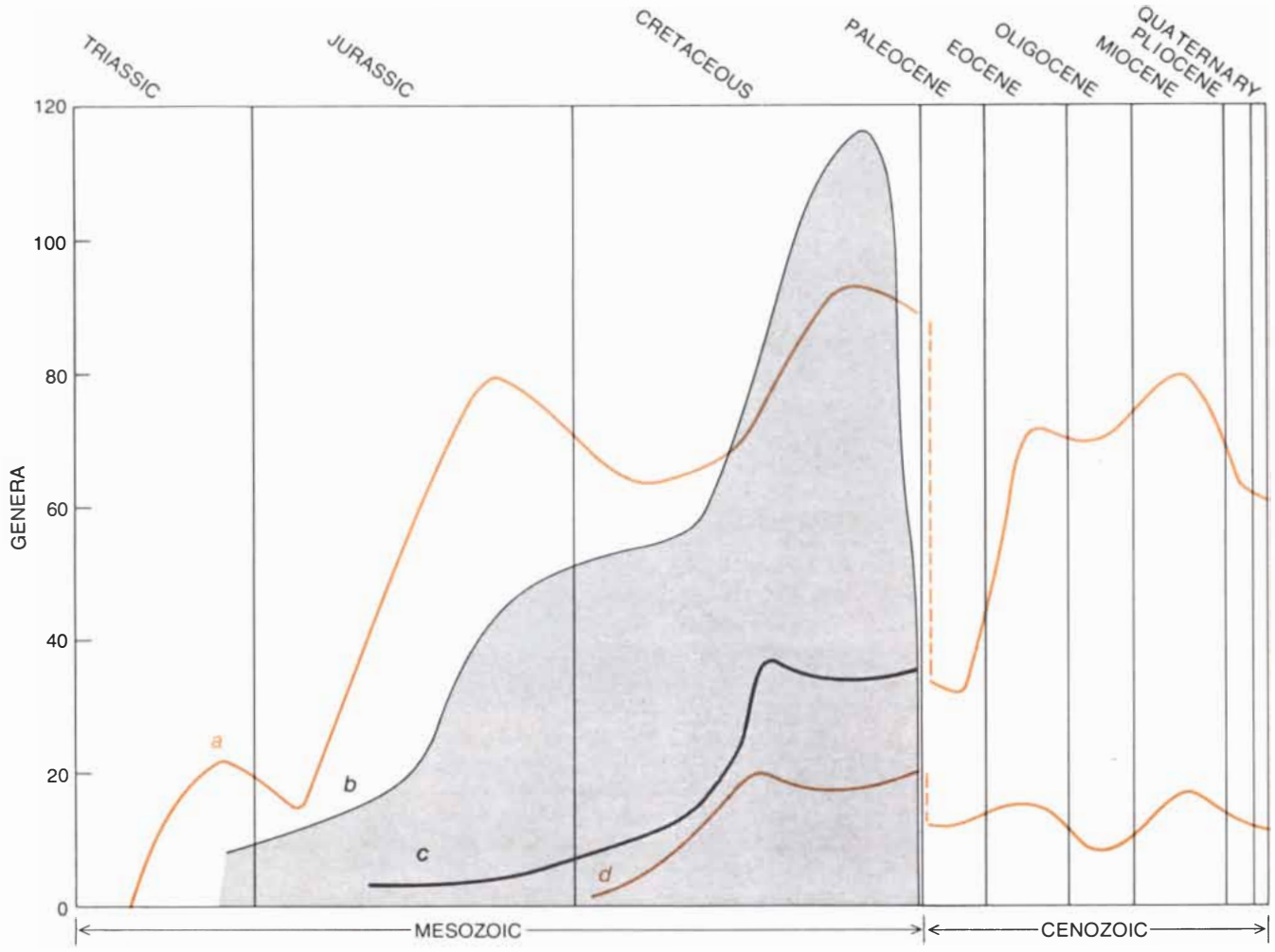
The environmental changes that devastated life at sea also took their toll of land animals. Perhaps the most spectacular instance of extinction ashore involved the group that had been the dominant higher animals during most of the Mesozoic era: the dinosaurs. Of the 115-odd genera of dinosaurs found in late Cretaceous fossil deposits, none survived the end of the period. The concurrent breakdown of so many varied communities of organisms clearly suggests a single common cause.

What was the source of this biological crisis? We have now come sufficiently close to the present to have a better grasp of the evidence. Throughout almost all of Mesozoic times life both on land and in the sea seems to have been remarkably cosmopolitan. A broad belt of equable climate extended widely in both directions from the Equator and there was no very evident segregation of organisms into climatic zones. The earth has always been predominantly a water world; the total land area may never have exceeded the present 30 percent of the planet's surface and has often been as little as 18 percent. In the late Cretaceous almost two-thirds of today's land area was submerged under shallow, continent-invading seas. It appears that during times of extensive inundation such as this one there was nothing like the blustery global circulation of air and strong ocean currents of today.

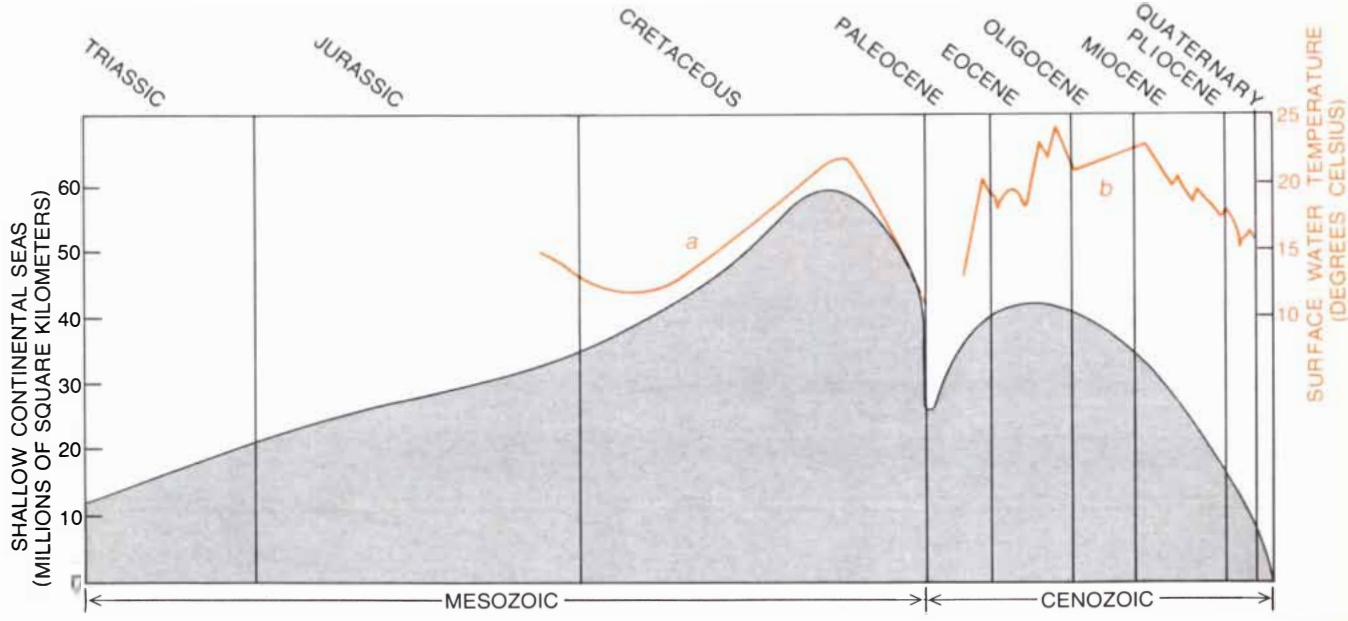
Paleoclimatology reveals the contrast between contemporary and Cretaceous conditions. By measuring the proportions of different isotopes of oxygen present in the carbonate of fossil foraminifera and mollusks it is possible to calculate the temperature of the water when the animals were alive. Today the temperature of deep-ocean water is about three degrees C. Cesare Emiliani of the University of Miami has shown that early in the Miocene epoch, some 20 million years ago, the bottom temperature of the deep ocean was about seven degrees C. He finds that in the Oligocene, 10 million years earlier, the temperature was about 11 degrees and that in late Cretaceous times, 75 million years ago, it was 14 degrees. He suggests that the onset of cooling may have been lethal to dinosaurs. In any event it is clear that cold bottom water has been accumulating in the deep-ocean basins since early in the Cenozoic era.

The accentuated seasonal oscillations in temperature and rainfall at the end of the Cretaceous period, a trend that evidently started when the shallow continental seas began to drain away into the deepening ocean basins, have been credited by Emiliani and also by Daniel I. Axelrod of the University of California at Davis and Harry P. Bailey of the University of California at Riverside with the simultaneous reduction in numbers or outright extinction of many other animal and plant species at that time. Genetically adapted to a remarkably equable world climate over millions of years, many of these Mesozoic organisms were ill-prepared for the extensive emergence of land from under the warm, shallow continental seas and the climate of accentuated seasonality that followed. Perhaps we should be surprised not that so many Cretaceous organisms died but that so many managed to adapt to the new conditions and thus survived.

Near the end of the Paleocene epoch, some 10 million years after the great



MESOZOIC INCREASE in the diversity of ocean and land animal genera reached a peak in the Cretaceous period. It was followed both by extinctions and by severe reductions in the number of genera that chanced to survive. At sea the explosively successful rudists (*c*) and ashore the long-dominant dinosaurs (*b*) became extinct. Corals (*a*) and globigerine foraminifera (*d*) fell in numbers.



SHRINKING OF SEAS that had invaded large continental areas (*gray*) began near the end of the Mesozoic era and accelerated in the Miocene epoch. Oxygen isotopes in belemnite (*a*) and foraminifera (*b*) skeletons show that the ocean has become progressively cooler since early in Miocene times (*color*). A relation seems to exist between a larger land area and greater seasonal extremes in climate.

Cretaceous collapse, a reef community sans rudists reappeared in the tropical seas. The following epoch, the Eocene, saw a new radiation of the scleractinian corals. Several genera that now appeared worldwide were unknown earlier in the fossil record; many of them are still living today.

The Cenozoic Decline

A sharp reduction in coral diversity that began in late Eocene times and lasted throughout the Oligocene epoch seems to reflect a continued increase in seasonality of climate and a substantial lowering of mean temperatures over large areas of the globe. Nonetheless, communities built around a bipartite association of corals and coralline algae continued to build extensive reefs in the Gulf of Mexico and the Caribbean area, in southern Europe and in Southeast Asia. Continued sea-floor spreading and deepening of the Atlantic basin, which enhanced the effectiveness of the Atlantic as a barrier to migrating corals, are evidenced in increasing differences between Caribbean and European coral species in late Oligocene fossil deposits.

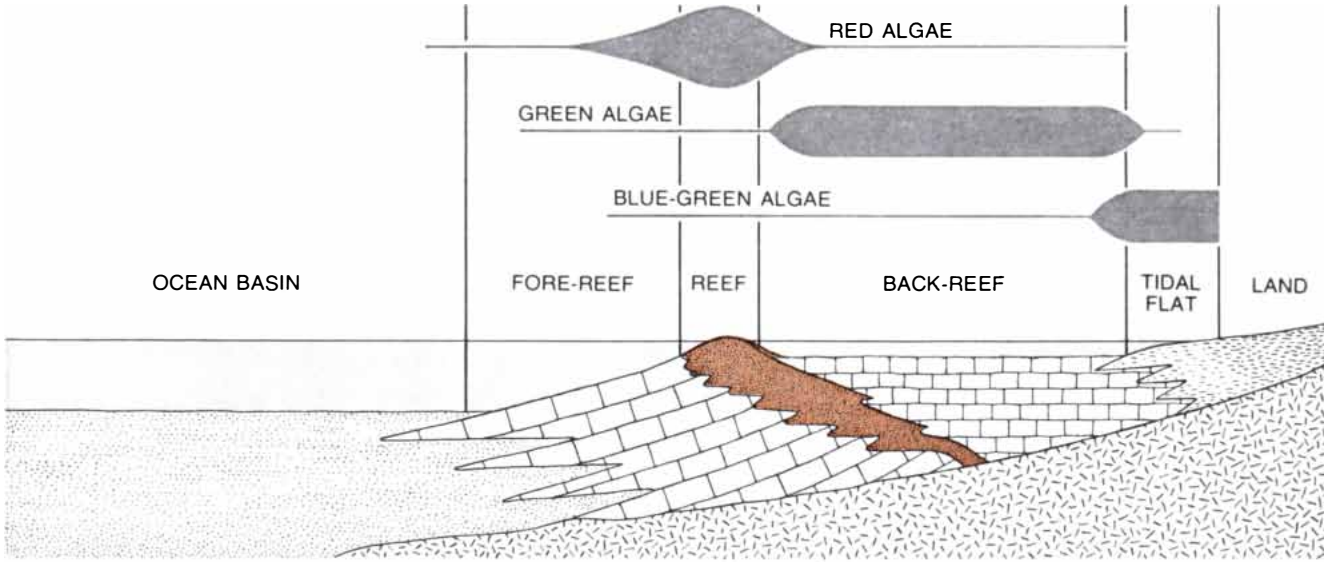
By now the earth had been free from continental glaciers for almost 200 million years, but a change in the making was to have profound consequences for world climate: the Antarctic ice cap was becoming established. Fossil plant remains and foraminifera from nearby deep-sea deposits both testify that even in early Miocene times the climate in much of the Antarctic continent was not much different from the blander days of the early Cenozoic, when palm trees grew from Alaska to Patagonia. At this time, moreover, Antarctica was some distance away from its present polar position. Nonetheless, mountain glaciers had begun to appear there millions of years earlier, in Eocene and Oligocene times. Sands of that age, produced by glacial streams and then rafted out to sea on shelf ice, are found in offshore deep-sea cores. The cooling trend was well established before the end of the Miocene epoch. In the Jones Mountains of western Antarctica lava flows of late Miocene age overlie consolidated glacial deposits and extensive areas of glacially scoured bedrock.

With the formation of the Antarctic ice cap some 15 to 20 million years ago a factor came into being that strongly influences world weather patterns to this day. The ice cap energizes the world weather system. In the broad reaches of open ocean surrounding Antarctica the surface water is aerated and cooled until it is too heavy to remain at the surface. The cold water sinks and spreads out along the sea floor, following the topography of the bottom. The result is a gravity circulation of cold water from Antarctica into the world's ocean basins, with a consequent lowering of the mean ocean temperature and cooling of the overlying atmosphere. The energetic interactions of the atmosphere above with the ocean and land below, in turn, strongly influence global wind patterns and worldwide weather. Today's climate is the product of a long cooling trend, marked by ever greater seasonal extremes; the trend became strongly accentuated when the Antarctic ice cap came into being late in the Miocene epoch.

This event and others in the Cenozoic era are faithfully recorded in terms of changes in the reef community. For example, in spite of the development of new barriers to the migration of reef organisms during the Mesozoic era, such as the Atlantic deep, the reef community had remained predominantly cosmopolitan up to the close of the Cretaceous. By Miocene times, however, what had once been essentially a single pantropical community was effectively divided into two distinct biogeographic provinces: the Indo-Pacific province in the Old World and the Atlantic province in the New World.

In the Old World the increasingly unfavorable climate had eliminated the reef community in European waters. It was during Miocene times, when Australia reached its present tropical position, that the Old World reef-builders first began to colonize the shallows of the Australian shelf; in terms of maximum diversity, the headquarters of the Indo-Pacific province today lies in the Australasian region where seasonal contrasts in water temperature are minimal.

Like the Atlantic deep, the deep waters of the eastern Pacific formed a generally effective barrier to the migration of reef organisms from the Indo-Pacific province into the hospitable tropical waters along the west coast of the Americas, principally around Panama. In Miocene times this Pacific coastal pocket was still connected to the reef-rich Caribbean, the headquarters of the At-



REEF ANATOMY centers on a rigid core (*color*) composed of the cemented skeletons of algae and corals. In this diagram growth of the core began at right and continued upward and outward. Most of the bulk of the reef consists of wide areas of stabilized detritus that are continuously being converted into the limestone that comprises the fore-reef and back-reef. The sandlike detritus is stabilized by the growth of the animals and plants in the reef community. The zones occupied by different algae are indicated here.

lantic province. Contact between the Atlantic province and the small Pacific enclave continued until the two areas were separated during the Pliocene epoch by the uplifting of the Isthmus of Panama.

The Pliocene saw a further contraction of the world tropics. The community of reef-builders gradually retreated to its present limits, generally south of 35 degrees north latitude and north of 32 degrees south latitude. Rather than serving as a center for new radiations, this tropical belt essentially became a haven. The epoch that followed, the Pleistocene, was marked by wide fluctuations in sea level and sharp alternations in climate that accompanied a protracted series of glacial advances and retreats. Oddly enough, the reef community was scarcely affected by such ups and downs; the main reasons for this apparent paradox seem to be that neither the total area of the deep tropical seas nor their surface water temperature underwent much change during the Pleistocene ice ages.

Is a fifth collapse in store for the most complex of ocean ecosystems? It would be foolhardy to respond with a flat yes or no. If, however, the past is prologue, the answer to a collateral question seems clear. The question is: Will a reef community in any event survive? The most significant lesson that geological history teaches about this complex association of organisms is that, in spite of the narrowness of its adaptation, it is remarkably hardy. At the end of the long interlude that followed each of four successive collapses, the reef community entered a new period of vigorous expansion. Moreover, without exception each revitalized aggregate of reef-builders included newcomers in its ranks.

The conclusion is inescapable that even during the times most unfavorable to the reef ecosystem the world's tropical oceans must have had substantial refuge areas. In these safe havens many of the threatened organisms managed to adapt and survive while others seem to have crossed some evolutionary threshold that had prevented their earlier appearance in the community. Today the Atlantic and Indo-Pacific provinces are two such refuge areas.

As long as the cooling trend that began in Cretaceous times does not entirely destroy these tropical refuge areas, one long-range conclusion seems firm. Any collapse of the present reef community will surely be followed by an eventual recovery. The oldest and most durable of the earth's ecosystems cannot easily be extirpated.

How some oil companies see it.

How we see it.

At American Oil, we look for everything that makes a product important to you.

We take a closer look at all of the conditions under which you use our industrial oil, so we can also take a better look at the oil. That's how we keep giving you newer and better products for your industry.

At American Oil, it's not just what we see . . . it's how we see that makes us important to you.

AMERICAN

American Oil Company

POLAROID AND THE EMERGENCE OF FYDEL JONES.

Fydel Jones is a Havasupai Indian. He lives on a tiny reservation deep in the Grand Canyon.

Preschool Havasupai kids spoke little or no English and their teacher, when he arrived, spoke no Havasupai. But a worker in the Head Start program thought of Polaroid Land cameras as a way to help bridge the gap.

It worked. The children got involved. They took pictures of familiar objects. The teacher identified the objects verbally and in writing. Vocabulary grew. Sentence structure developed.

And the small Havasupai, having learned English, were ready for school in the world outside the Canyon.

We present this story not as an end in itself, but rather as an example of a phenomenon...the proliferating uses of the Polaroid Land camera as an educational tool. Visual aids in education are not new. But the particular and potent advantages offered by instant photography are becoming more and more widely appreciated.

Helping children to learn involves psychological and emotional factors. To create the necessary motivation, interest must be aroused and maintained. Fears must be overcome. And a feeling of accomplishment is needed to generate continuing interest and activity.

The Polaroid Land camera is simple to use, and because you can re-shoot pictures on the spot, corrections are easy to make. So there is no fear of failure. And results are immediate; so there is involvement and a sense of success from the outset. There is even the higher accomplishment of bringing out creative self-expression.

The reasons for the rapid spread of instant photography as an educational tool are therefore clear. But in certain applications its benefits are even more dramatic. For example, in schools for the profoundly deaf, it helps children learn to lip read; and in "inner city" programs, it is used where children urgently need a feeling of involvement.

Two organizations, The Environmental Studies Project, of Boulder, Colorado, and Education Development Center Inc., of Newton, Massachusetts, have gone even further. They have developed classroom materials and entire programs for teachers, with instant photography as an essential component.

Educational results can not always be precisely measured. Nor are they always as clearly evident as in the case of the Havasupai. But there has been an impressive volume of reports from educators in the field, at every level, and in conventional as well as specialized schools. And if the sum of their comments is a valid yardstick, we have reason to be proud of the part instant photography is playing. Which is, through aroused and sustained motivation, to help potential intellect to truly "emerge."

Polaroid Corporation

Temperature Control in Flying Moths

The flight muscles of certain insects cannot function properly until they have been warmed up by a process that resembles shivering. The system also keeps the insect from overheating

by Bernd Heinrich and George A. Bartholomew

When a housefly takes off in the kitchen, it does so almost instantaneously. In cold surroundings the fly is torpid and cannot take off at all. The housefly, like most insects, is a cold-blooded animal, and its "flight motor" cannot function efficiently at low temperatures. Certain insects, however, are able to fly in low air temperatures by warming up their flight motor by a process related to shivering in vertebrates. The lower the external temperature is, the longer the preflight warm-up takes; at times it goes on for 15 minutes or more.

We have investigated this phenomenon with various species of sphinx moth, some of which reach the size of hummingbirds. It turns out that biological engineering of remarkable virtuosity is involved in the process whereby these insects develop and maintain the necessary temperature in their flight muscles. The phenomenon represents an independent evolutionary creation of an elaborate pattern of body-temperature regulation that differs in interesting ways from the patterns evolved by terrestrial vertebrates.

Any organism's expenditure of energy involves an elaborate web of chemical reactions, the greater part of which are sensitive to temperature. Indeed, the rates of most chemical reactions double with each increase of 10 degrees Celsius in temperature, a generalization first made by the great Dutch chemist Jacobus Hendricus van't Hoff. The importance of environmental temperature to poikilothermic, or cold-blooded, animals is therefore obvious: their body temperature varies directly with the environmental temperature. The situation is less obvious with homeothermic, or warm-blooded, animals, in which body temperature is virtually independent of environmental temperature over a wide range of conditions, although the environmental temperature is still important. Maintaining homeothermy is energetically expensive. Even under the least demanding circumstances, birds and mammals devote 90 percent or more of their total energy metabolism to the regulation of body temperature, and for many species the energy cost of homeothermy increases markedly as the environmental temperature decreases.

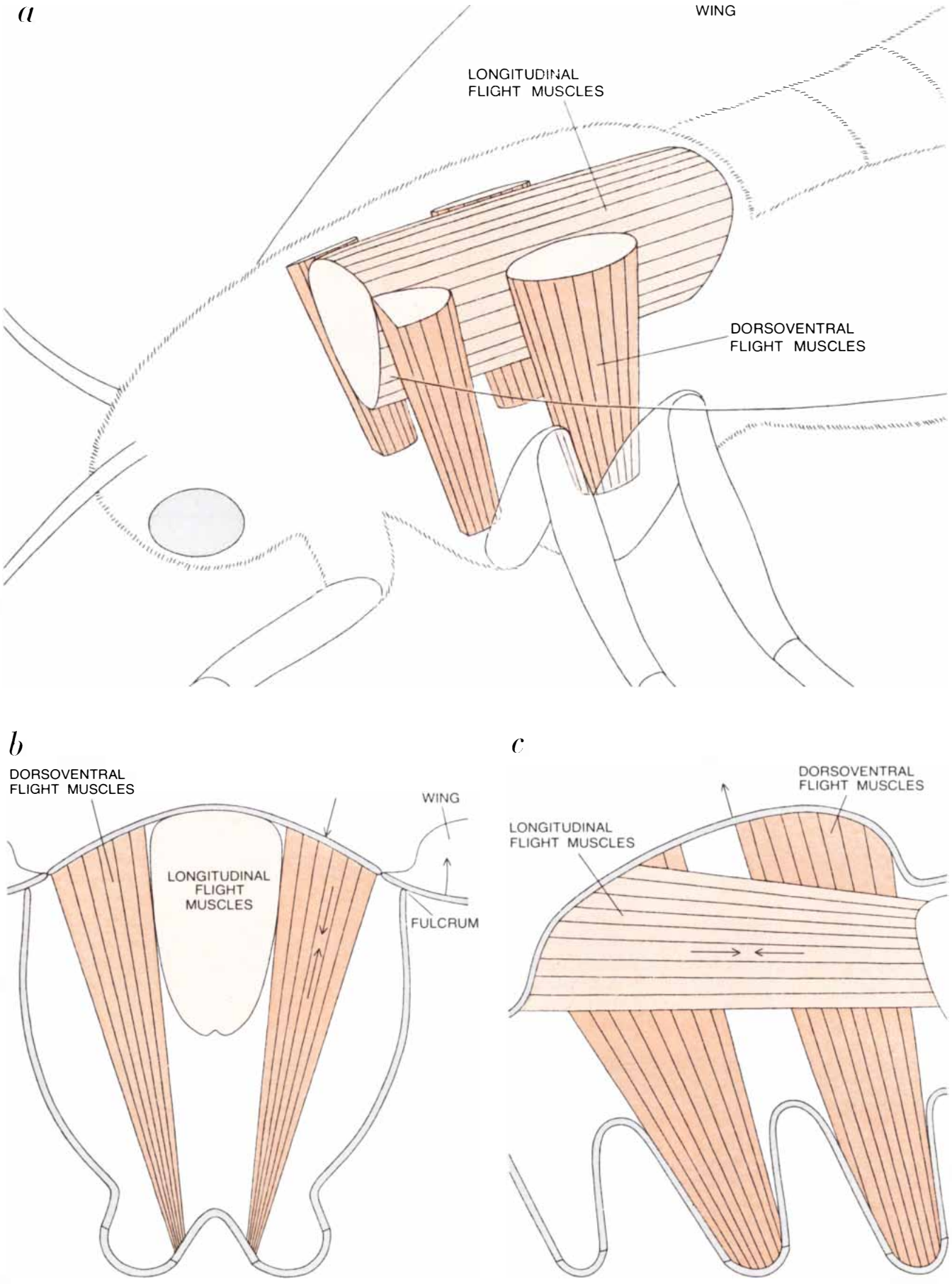
Most small invertebrate animals have such high rates of heat transfer between themselves and the environment that their energy metabolism contributes little or nothing to the heat content of the body. Such organisms are called ectothermic because the heat that determines their body temperature comes almost entirely from convective, conductive and radiative exchanges with the environment. In another category are the endothermic animals, in which the heat that determines body temperature is derived largely from the animal's own energy metabolism. In general these animals are larger than the invertebrates, and many of them are well insulated. These animals have high rates of metabolic heat production and relatively low rates of heat exchange with the environment.

Another term, heterothermy, characterizes animals that can produce enough heat to sustain an elevated body temperature but that become essentially ectothermic when they are inactive or dormant. In this group are such animals as certain hummingbirds, bats, ground squirrels and hedgehogs. Some of these animals have daily periods of dormancy and some have seasonal periods of hibernation or estivation. Even though the usage is somewhat unconventional, the term heterothermic is equally applicable to other groups, such as some of the insects, that are endothermic only during periods of activity.

Although ectotherms obtain the heat that determines their body temperature from the environment, it is important to recognize that some of them can nonetheless control their body temperature with high precision through their behavior. For example, a desert lizard can maintain its body temperature within a degree or so of 35 degrees C. for many hours at a time, even though the environmental temperature (of the air and the ground) fluctuates through a range of 20 to 30 degrees. The lizard accomplishes the feat by moving from place to place, exploiting both solar radiation and the thermal variability of the environment.

Even with this capability an ectotherm is inevitably a partial prisoner of its thermal environment. An endotherm, on the other hand, is relatively independent of environmental temperature and so is relatively free of this major constraint. The selective advantage of having a body temperature determined by heat produced metabolically is demonstrated by the fact that endothermy has evolved independently several times. Moreover, one finds at least partial endothermy in certain animals that have not evolved full endothermy. For example, the female of the Indian python can

SPHINX MOTH of the species *Manduca sexta*, which was studied by the authors, is shown from above in the photograph on the opposite page in its characteristic hovering flight, which resembles the hovering flight of a hummingbird. The nocturnal moth is cold-blooded when at rest. Under most circumstances the temperature is not high enough for the flight muscles to function well. The moth solves this problem by warming up its flight muscles with metabolic heat.



FLIGHT MUSCLES of a sphinx moth are depicted schematically (*a*). The dorsoventral ones are involved in the upstroke, the longitudinal ones in the downstroke. The body wall serves as a fulcrum for the wing (*b*), so that when the dorsoventral muscles contract, the wing is raised. Contraction of the longitudinal muscles (*c*) flexes the upper side of the insect, forcing the wings downward.

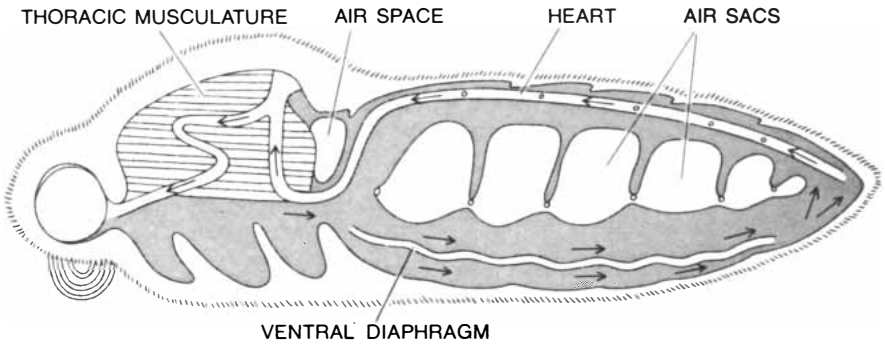
produce enough heat metabolically to keep its body temperature seven degrees C. or more above the ambient temperature while it is incubating eggs. Large, fast-swimming fishes of the mackerel family, such as tuna, can keep their central body temperature near 32 degrees C. while swimming in water of 20 degrees. Mako and porbeagle sharks have a similar capacity.

All these animals are large, and their ability to achieve partial endothermy can be rationalized in terms of thermal inertia, favorable surface-to-volume ratio and other conditions that keep heat transfer low. Insects, however, are small and cannot take advantage of such conditions. How, then, does one explain the partial endothermy of insects?

It has been known for more than a century that the temperature of an insect in flight can be higher than the environmental temperature. In 1928 Heinz Dotterweich of the University of Kiel showed that certain sphinx moths elevate the temperature of their thorax (the front part of the body) before flight, and in the early 1940's Torkel Weis-Fogh (now at the University of Cambridge) and Erik Zeuthen of the University of Copenhagen reported similar observations on members of the orders Coleoptera (beetles) and Hymenoptera (bees and wasps). In 1965 James E. Heath of the University of Illinois and Phillip A. Adams of California State College at Fullerton showed that sphinx moths can maintain high and stable body temperatures while flying.

The question of how insects generate heat and regulate body temperature has been investigated in detail only in large moths of the families Saturniidae and Sphingidae, which weigh from one gram to several grams. It is known, however, that bumblebees weighing as little as 100 milligrams can maintain thoracic temperatures 20 degrees C. or more above the ambient temperature by means of internally produced heat. Since our own studies have been concerned primarily with sphinx moths, we shall confine our discussion mainly to this group.

The larger sphinx moths weigh from two to six grams, thus overlapping the size range of the smaller hummingbirds, with which they show striking evolutionary convergences in behavior, physiology and body form. Indeed, the adult moths feed on the nectar of flowers while hovering, as hummingbirds do, and while in flight can easily be confused with hummingbirds. Although sphinx moths are found throughout the world except in the Antarctic, most species are nocturnal and therefore often escape notice. The larvae of certain species are better known because they are agricultural pests; the caterpillars of *Manduca sexta* (a species that we studied in the adult form in a number of our experiments) are the hornworms that feed on tobacco and tomato plants.



CIRCULATORY SYSTEM of a sphinx moth is portrayed in this longitudinal section of the insect's body. Pulses of the long heart usually travel from back to front. From the aorta the blood flows into the abdomen, where its movement is assisted by the ventral diaphragm. The air space between the thorax and the abdomen helps to retain heat in the thorax. If the thorax gets too warm, the heart beats faster and increases blood flow. Blood carries the excess heat into the otherwise unwarmed abdomen, where it is dissipated to the environment.

The "motor" that drives the wings of a sphinx moth consists of a complex of muscles in the thorax [*see illustration on opposite page*]. The flight muscles of insects are metabolically the most active tissue known, and for *M. sexta* to drive its wings at the required rate of 25 to 30 beats per second during flight requires a prodigious expenditure of energy. Moreover, the flight muscles of the population of *M. sexta* from the Mojave Desert that we studied cannot achieve the power output required for effective hovering flight unless they are at a temperature of 35 to 38 degrees C. Nevertheless, these sphinx moths can be seen flying when the air temperature is as low as 10 degrees C.

When the moths are at rest, they are ectothermic; the body temperature is almost indistinguishable from the environmental temperature. Since the moths are nocturnal and the environmental temperature at night in the Mojave Desert is almost never high enough to bring their flight muscles to operating temperature, they would be permanently grounded if they were exclusively ectothermic, with a body temperature depending entirely on heat from external sources. They are able to fly because they have a capacity for partial endothermy, which they employ to warm up their flight muscles before they try to take off.

The fuel for the flight muscles of moths is primarily fat. When it is oxidized, it produces carbon dioxide, water and large quantities of heat. The heat raises the temperature of the flight muscles, and the increased temperature (in conformance with van't Hoff's generalization) enables the muscle tissues to burn the fuel faster, which in turn produces more heat. This process continues until the metabolic rate becomes high enough to generate the enormous output of work needed to support the heavy moth in flight.

In its warm-up the sphinx moth assumes an elevated stance and vibrates its wings through a small arc. The upstroke and downstroke wing muscles are activated by their respective nerves in approximate synchrony instead of alternately as in flight, which is why the warm-up resembles the shivering of birds and mammals. If the air temperature is above some 10 degrees C., the muscle contractions produce heat more rapidly than it is lost to the environment, so that the temperature of the flight muscles begins to rise. The hotter the flight motor becomes, the higher its rate of heat production is. The progressive rise in thoracic temperature increases not only the rate of heat production but also the rate of heat loss to the environment. As a result the relation between heat production and heat loss becomes constant soon after the insect starts warming up, so that during most of the warm-up the thoracic temperature increases linearly rather than exponentially. Throughout the warm-up the temperature of the abdomen does not change, remaining close to the external temperature.

The duration of the preflight warm-up depends on the temperature of the environment. The reason is that the rate of increase in thoracic temperature increases directly with the ambient tem-

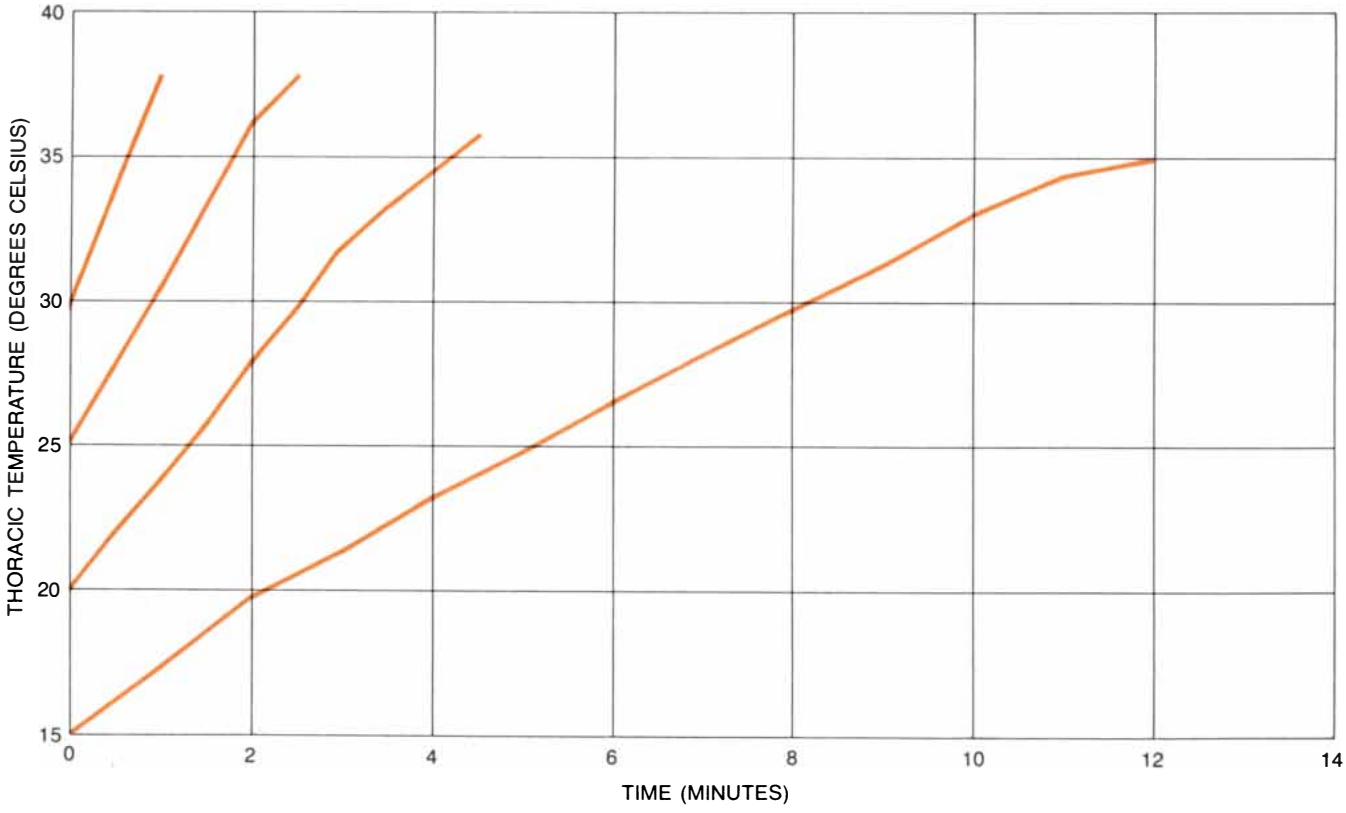
perature. Moreover, the higher the ambient temperature, the smaller the increment required to reach flight temperature. At an environmental temperature of 30 degrees C. the sphinx moth can get its flight muscles to takeoff temperature in about a minute. When the external temperature is at 15 degrees, the insect may need 15 minutes of warm-up. The relation between ambient temperature and length of warm-up is to be expected on physical grounds. What is most remarkable about the procedure is that natural selection has produced a system whereby internally produced heat increases the temperature of part of the body so effectively in an animal of such a disadvantageous surface-to-volume ratio.

How is such thermoregulatory virtuosity achieved? A number of factors other than the high metabolic rate of the flight motor can be identified. One is that the thorax is insulated from the abdomen (the rear part of the body) by a large air space, so that conductive heat transfer from the thorax to the abdomen is minimized. Another factor is that the rate of convective heat loss to the air is minimized by a dense insulating cover of long scales on the thorax. In contrast, the scales on the abdomen are short and appear to have relatively little effect on heat transfer. One can test the insulative effect of the thoracic scales by measuring the cooling rate of the thorax with the scales present and then with the scales removed. The rate is appreciably higher when the scales are missing.

These findings are helpful but do not carry the investigation very far. A fuller understanding of the physiology of the sphinx moth's thermoregulation requires more complex procedures and an examination of the levels of body temperature maintained during flight. One soon perceives that the mechanism for warm-up is only part of what is required. The system must also contain a means of keeping the thoracic temperature from rising too high.

It is noteworthy that the minimum thoracic temperature of 36 to 38 degrees C. for efficient hovering flight by *M. sexta* is in the same range as the body temperatures of most vertebrate homeotherms. Only songbirds have higher temperatures (40 to 42 degrees), and during hovering flight the moth's thoracic temperature reaches even that level. Here is where the problem of keeping the temperature from rising too high is encountered. In insects as in birds and mammals even a small increase in body temperature above the level characteristic of normal activity can result in serious and sometimes irreparable tissue damage through the denaturation of enzymes, coagulation of proteins and other difficulties. Hence the body temperature during flight must be regulated so that it remains high enough for hovering (40 degrees C.) but below the level of tissue damage (46 to 47 degrees in this species).

As a first step in investigating the dynamics of temperature control in *M. sexta* we examined the relation of thoracic temperature to air temperature during flight. Moths were allowed to fly freely in a temperature-controlled room with the temperature ranging from 12 to 40 degrees C. After a moth had flown long enough to reach a constant body temperature (at least two minutes) it was caught and its body temperature was quickly measured by inserting into the thorax a tiny thermocouple connected to a recording potentiometer. The moths flew readily in air temperatures from 12 to 35 degrees but were reluctant to fly at temperatures above 35 and would not fly for more than one minute at 40 degrees. Between 20 and 30 degrees thoracic temperatures during uninterrupted flight were virtually independent of air temperature, remaining near 40 or 41 degrees. Thoracic temperatures were

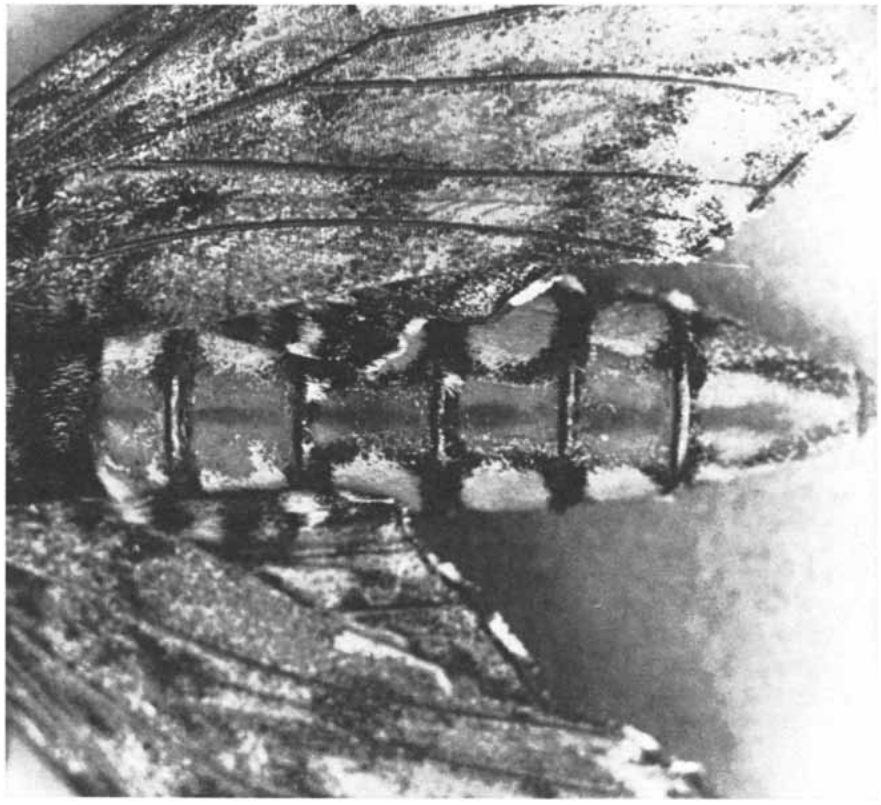


EFFECT OF WARM-UP on the thoracic temperatures of a number of sphinx moths is charted for four different levels of air temperature. Reading from the bottom, the air temperatures were 15, 20, 25 and 30 degrees Celsius. In other words, it took the moths about 12 minutes to bring the flight muscles to operating temperature at an outside temperature of 15 degrees, and one minute at 30 degrees.

lower at air temperatures below 20 degrees and higher at air temperatures above 35 degrees.

Heath and his co-workers at the University of Illinois have obtained similar results, although they report thoracic temperatures several degrees lower than the ones we found. It is not yet clear whether the differing results are due to different experimental procedures or to physiological differences between the desert population of *M. sexta* and the Middle Western population.

How can the thoracic temperature of a flying moth be so independent of the air temperature? The rate of heat loss of an object is directly related to the difference between the temperature of the object and the environmental temperature. Therefore if the thoracic temperature holds steady at 40 degrees C., the rate of heat loss from a moth flying at 15 degrees must be about twice the loss from a moth flying at 28 degrees. Plainly the moth must be able to vary either heat production or heat loss or both in order to maintain a thoracic temperature that is largely independent of the air temperature.



HEART OF SPHINX MOTH is the slender dark line running up the middle of the insect's back. It is visible in this photograph because scales have been removed from the moth's abdomen, so that the heart shows through the transparent tergites, or plates, of the body.

We tested in two ways the question of whether or not heat production is varied. One procedure involved building a flight mill to which a moth could be tethered for controlled flight. A moth thus tethered is partly supported by the rotating arm of the mill, so that the loading of the insect's wings is reduced. Since under these conditions less energy is required for flying, less heat should be produced than during free flight and thoracic temperature should be lower. If significant quantities of heat are generated specifically for temperature regulation, however, the thoracic temperature of a moth on the mill should be as high as it is in free flight.

We found the thoracic temperatures of the moths during captive flight were much lower than they were during free flight. In captive flight the thoracic temperature varied almost directly with the air temperature. We interpret these findings to be the result of diminished heat production associated with reduced wing loading, and we conclude that under the experimental conditions little or no heat was produced specifically for temperature regulation, the heat being merely a by-product of the flight metabolism.

The experiments with the flight mill yielded additional insights on the thermoregulatory mechanisms of *M. sexta*. When a moth flew on the mill, its thoracic temperature varied with the air temperature until the thoracic temperature approached 40 degrees C. Above that level it tended to stabilize, and the difference in temperature between air and thorax diminished conspicuously. Apparently the thoracic temperature was not regulated until it reached a level where the danger of overheating became evident.

Our second method of determining whether or not heat production during flight might be varied for thermoregulatory purposes involved measuring the rates of oxygen consumption of moths in free flight. Because of the ability of the moths to hover, they could be flown inside large glass containers. In this way we could determine, with the aid of a paramagnetic oxygen analyzer, how much oxygen a moth consumed during measured periods of uninterrupted flight over a range of ambient temperatures. We found that a 1.5-gram sphinx moth consumed some 67 to 75 milliliters of oxygen per hour and that this consumption, which corresponds to a heat production of five to six calories per minute, was independent of the ambient temperature. From the latter finding we infer that heat production during uninterrupted flight is not varied for thermoregulatory purposes. The findings from the flight-mill experiments and from the oxygen-measuring experiments were in accord.

The significance of the data that we had obtained up to this point was that a sphinx moth must be able to control the rate of heat loss from its thorax to the environment. The obvious starting point in a search for a mechanism of control was the circulatory system. The circulatory system of insects is open-ended, contrasting sharply with the closed system of vertebrates. The hemolymph (blood) of an insect moves freely through open spaces in the body, bathing the internal organs.

In sphinx moths and other sphingids the tubular heart has a larger diameter than the heart of most insects. The heart extends the length of the dorsal part of the abdomen and into the thorax. If the scales of the abdomen are removed along the insect's back, the heart is clearly visible through the transparent plates under the scales. The pulsations of the heart can be seen as they proceed from rear to front, carrying blood from the abdomen to the thorax. The heart continues into the thorax as a large vessel, the aorta, which loops through the flight muscles.

We investigated the role of the circulatory system in temperature control

by tying the heart at the point where it leaves the abdomen and enters the thorax. In this way the circulation of blood between the abdomen and the thorax was eliminated or greatly reduced. Since the primary limiting factor in insect flight is oxygen and insects obtain oxygen through a set of tubes that is separate from the circulatory system, the moths could continue to fly even though little or no blood was reaching their flight muscles.

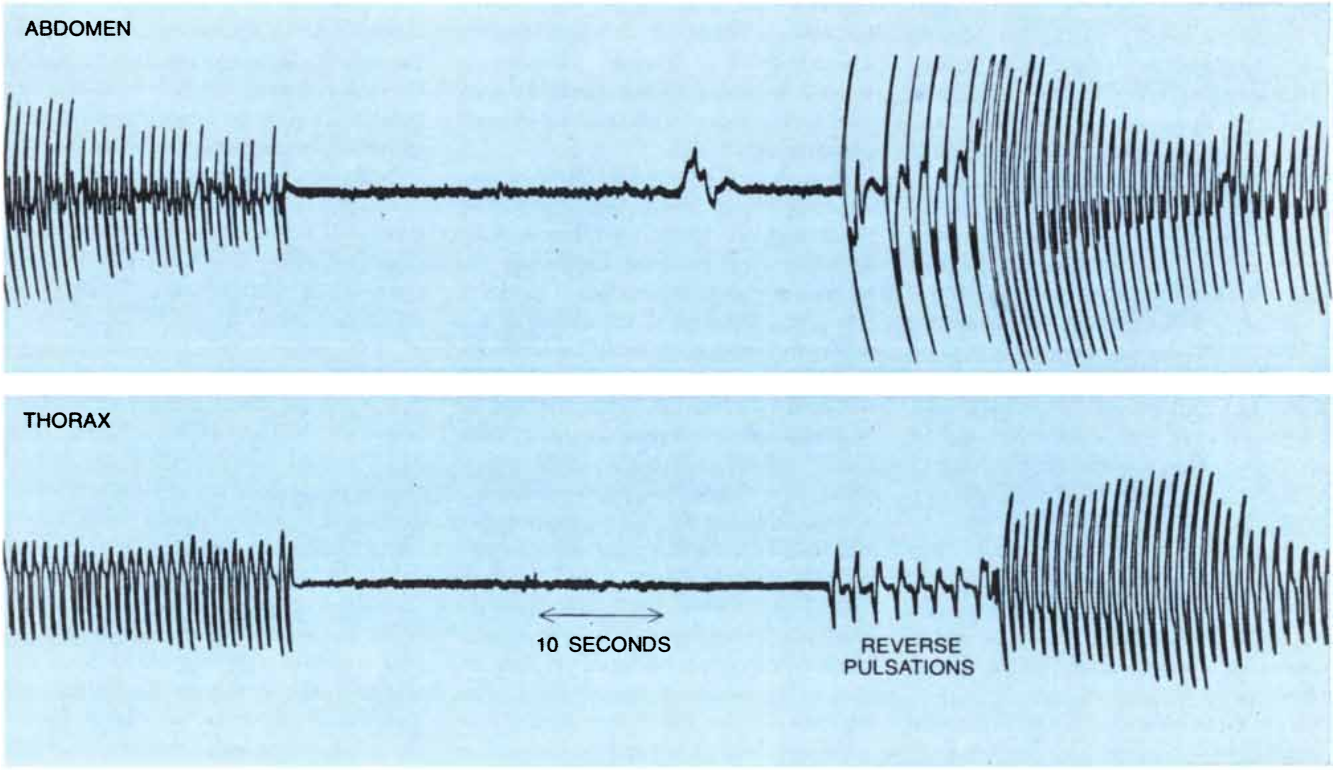
During the preflight warm-up the rate and extent of increase of thoracic temperature in the insects with tied hearts was about the same as it was in insects with intact circulatory systems. What differed significantly was the ability of the moths with tied hearts to lose heat. After an operation to tie the heart, flying uninterruptedly even at temperatures as low as 23 degrees C. caused a sharp rise in thoracic temperature above the normally regulated level. Moreover, such moths were apparently unable to modulate the rate of loss of heat while flying. Their thoracic temperature was about 24 degrees higher than the air temperature and varied directly with it. The moths would fly only briefly before alighting, and until they cooled off they could not be induced to fly again.

It seemed clear that the refusal of these moths to engage in continuous flight at air temperatures above 23 degrees C. was due to overheating. The conclusion was reinforced when we removed the insulating layer of scales on the thorax, thereby increasing the rate of heat loss. Such moths flew readily and continuously at temperatures above 23 degrees, but they could remain in free flight only over the narrow range of air temperatures at which the rate of heat production by the flight motor and the uncontrolled rate of heat loss happened to result in the appropriate thoracic temperature.

These findings suggested that circulating blood in normal moths prevents overheating of the thorax by altering the rate at which heat is lost to the environment, thereby maintaining the temperature of the flight muscles within the range required for effective flight. The next question was how the control machinery works. The simplest situation would be for the circulatory system to function as an automatic regulator, with the rate of pulsation of the heart (and hence the rate of circulation of blood between the thorax and the abdomen) being determined by the temperature of the heart in the abdomen. If heat began to accumulate in the abdomen, the temperature of the heart would rise and, in conformity with the van't Hoff generalization, the heartbeat would accelerate. Blood would circulate more rapidly between the thorax and the abdomen and the rate of heat transfer from the thorax to the abdomen would automatically increase. Since the loss of heat from the abdomen to the environment would vary directly with the difference in temperature between them, the entire system would be self-regulating.

A more complex pattern, which would nonetheless be equally feasible from an engineering viewpoint, would be to have a thermostat in the flight motor controlling the rate of the heart's pulsations in the abdomen. We undertook to explore these alternatives by varying the temperature of the thorax while holding the temperature of the abdomen constant and vice versa. The means of heating was a beam of light. Body temperature and the rate and amplitude of heartbeat were recorded.

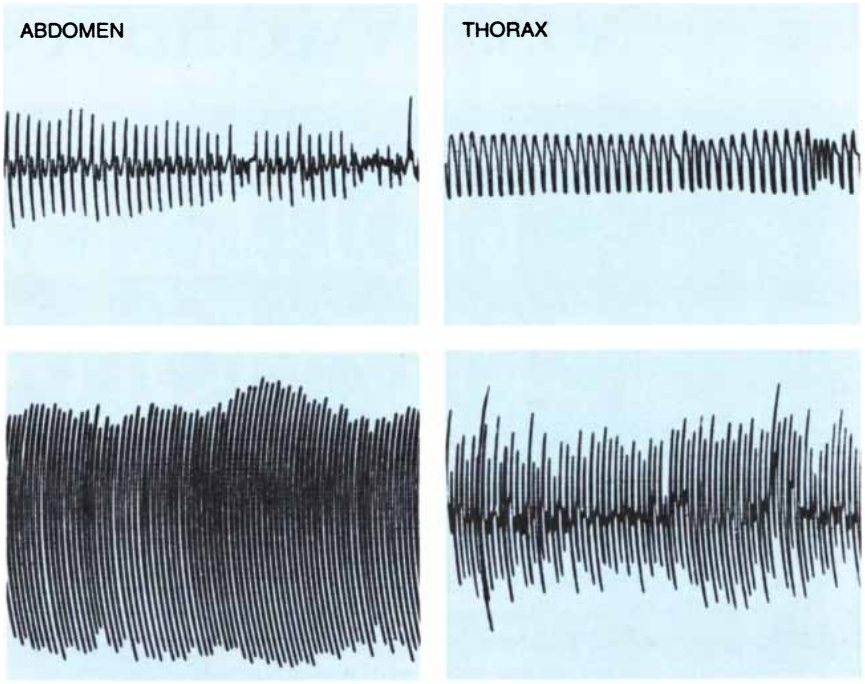
When the moths were at rest, the thorax and the abdomen were at about the same temperature as the air. The pulsations of the heart varied in rate and amplitude. Occasionally they reversed direction, traveling from front to rear. When heat was applied to the thorax but not to the abdomen, the thoracic temperature increased markedly; the abdominal temperature remained unchanged for a time. In the abdomen the pulsations of the heart increased in rate



HEARTBEATS of a resting moth were recorded electrically by means of an impedance converter to establish a basis for comparison of the rate and amplitude resulting when either the thorax or the abdomen of an insect was heated artificially in order to determine the mechanism for the dissipation of excess heat. The traces show heartbeat in the abdomen (*top*) and in the thorax (*bottom*).

and amplitude and traveled only from back to front. The pulsations of the heart were always at the same frequency in both the thorax and the abdomen even when the temperatures of the two regions differed by as much as 20 degrees.

Under these conditions the temperature of the thorax increased rapidly until it reached about 44 degrees C. There it stabilized, remaining below 44. Immediately after the thoracic temperature began to stabilize, the temperature of the abdomen began to rise sharply, even though heat was not applied to it. These data suggest that the rate and amplitude of pulsation in the abdominal heart are responsive to the temperature in the thorax and that overheating of the flight motor is prevented by a faster flow of blood, which rapidly transports excess heat to the abdomen, where it is dissipated to the air through the thinly insulated body wall.



ARTIFICIAL HEATING of a moth's thorax produced these effects on heartbeat. In each vertical set of two records the top tracing shows the heartbeat in the abdomen and the thorax before heating, when both abdominal and thoracic temperature were 25 degrees C. In the bottom tracings, made after heating, the thoracic temperature was 41 and the abdominal temperature 29. Thoracic temperature stayed near 41 and abdominal temperature rose, indicating that thoracic temperature control is achieved by carrying excess heat to abdomen.

We have conducted preliminary experiments that indicate the general nature of the mode of control of the heart rate. When the ventral nerve cord was cut behind the second abdominal segment, heating of the thorax did not evoke changes in the heartbeat. Moreover, the temperature of the thorax did not stabilize but continued to rise. Abdominal temperature increased by only two or three degrees rather than 10 or more, which was the case when the nerve cord was intact. Although the experiment was rather crude, it strongly suggests that the sensor controlling the heart rate and hence the dissipation of heat from the flight motor is in the thorax and maintains its control through nerve connections.

The thermoregulatory system of the sphinx moth shows both intriguing convergences with and fundamental differences from the systems evolved by birds and mammals. In birds and mammals the blood transports respiratory gases in addition to nutrients, hormones, metabolites and heat. Birds and mammals can control heat loss by such measures as elevating or depressing hair or feathers and contracting or dilating blood vessels, particularly in the limbs. Moths have no motor control over their insulating cover and do not have small blood vessels that can serve as local sites for heat exchange. Blood flow to the limbs is comparatively slight. Notwithstanding the differences in morphology, the increase in body temperature during the preflight warm-up of large moths is strikingly similar to the pattern observed in birds and mammals during arousal from the low-temperature state associated with hibernation, estivation and daily torpor.

It has long been known that small heterothermic animals warm up from a state of dormancy more rapidly than large ones. The same is true of heterothermic birds. The generalization that warm-up rate is inversely related to body weight can be extended to insects as a group, although most of the insects for which data are available appear to warm up somewhat faster than one would predict on the basis of data from birds and mammals. As far as is now known insects warm up only the thorax, whereas in birds and mammals the entire body is involved.

In any event it appears that at an air temperature of between 20 and 25 degrees C. an insect weighing 100 milligrams warms up at about 10 degrees per minute, whereas a bird or a mammal weighing 100 grams has a warm-up rate of about .3 degree per minute. Therefore an insect can pass through many cycles of heating and cooling in a day, whereas few mammals or birds go through more than one daily cycle. Indeed, daily torpor is known only in birds and mammals weighing less than 100 grams; heavier animals have warm-up rates that are too slow to fit into a daily cycle.

Many insects are too small to be able to regulate body temperature by producing heat metabolically. The large surface-to-volume ratio of such insects causes heat to be lost to the environment so rapidly that the temperature of the body undergoes little change. The same ratio, however, facilitates a rapid gain of heat from the environment, so that many insects obtain heat from the sun rather than through metabolic activity. Basking has been observed in cicadas, grasshoppers and flies; it is particularly familiar in butterflies. Such insects can be active throughout the day without incurring any added metabolic expense for regulating temperature. In contrast, nocturnal moths such as those we have studied, which rely on metabolic heat, usually have only brief periods of activity.

The findings that we and others have obtained show that the details of the mechanisms of temperature regulation in moths and other insects are intimately related to the structure of the insects and to many features of their adaptation to the environment. The present knowledge of temperature regulation in insects, however, is meager, having been obtained almost entirely from a few large species on which measurements can be made readily. The territory for investigation is wide, and much of it remains unexplored.

NONSTANDARD ANALYSIS

This mathematical theory has restored infinitesimals to good standing. They had been employed since antiquity, but often with doubts, to solve such problems as finding a circle's area

by Martin Davis and Reuben Hersh

Nonstandard analysis, a new branch of mathematics invented 10 years ago by the logician Abraham Robinson, marks a new stage of development in several famous and ancient paradoxes. Robinson, now at Yale University, has revived the notion of the "infinitesimal"—a number that is infinitely small yet greater than zero. This concept has roots stretching back into antiquity. To traditional, or "standard," analysis it seemed blatantly self-contradictory. Yet it has been an important tool in mechanics and geometry from at least the time of Archimedes.

In the 19th century infinitesimals were driven out of mathematics once and for all, or so it seemed. To meet the demands of logic the infinitesimal calculus of Isaac Newton and Gottfried Wilhelm von Leibniz was reformulated by Karl Weierstrass without infinitesimals. Yet today it is mathematical logic, in its contemporary sophistication and power, that has revived the infinitesimal and made it acceptable again. Robinson has in a sense vindicated the reckless abandon of 18th-century mathematics against the strait-laced rigor of the 19th century, adding a new chapter in the never ending war between the finite and the infinite, the continuous and the discontinuous.

In the controversies over the infinitesimal that accompanied the development of the calculus, Euclid's geometry was the standard against which the moderns were measured. In Euclid both the infinite and the infinitesimal are deliberately excluded. We read in Euclid that a point is that which has position but no magnitude. This definition has been called meaningless, but perhaps it is just a pledge not to use infinitesimal arguments. This was a rejection of earlier concepts in Greek thought. The atomism of Democritus had been meant to refer not only to matter but also to time and space. But then the arguments of Zeno had made untenable the notion of time as a row of successive instants, or the line as a row of successive "indivisibles." Aristotle, the founder of systematic logic, banished the infinitely large or small from geometry.

Here is a typical example of the use of infinitesimal arguments in geometry:

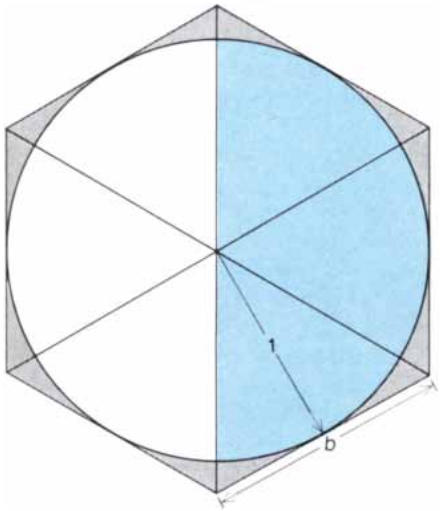
"We wish to find the relation between the area of a circle and its circumference. For simplicity we suppose that the radius of the circle is 1. Now, the circle can be thought of as composed of infinitely many straight-line segments, all equal to each other and infinitely short. The circle is then the sum of infinitesimal triangles, all of which have altitude 1. For a triangle the area is half the base times the altitude. Therefore the sum of the areas of the triangles is half the sum of the bases. But the sum of the areas of the triangles is the area of the circle, and the sum of the bases of the triangles is its circumference. Therefore the area of the circle of radius 1 is equal to one half its circumference."

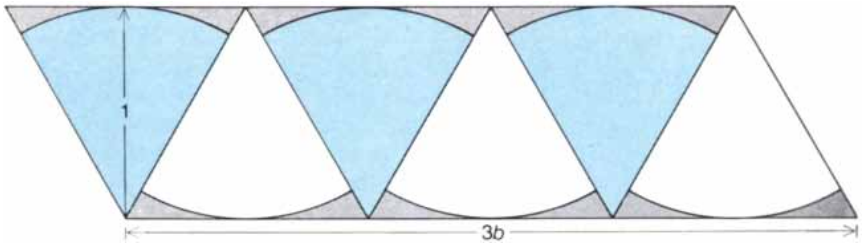
This argument, which Euclid would have rejected, was published in the 15th century by Nicholas of Cusa. The conclusion is of course true, but objections to the argument are not hard to find. The notion of a triangle with an infinitely small base is elusive, to say the least. Surely the base of a triangle must have length either zero or greater than zero. If it is zero, then the area is zero, and no matter how many terms we add we can get nothing but zero. On the other hand, if it is greater than zero, no matter how small, we will get an infinitely great sum if we add infinitely many terms. In neither case can we get a circle of finite circumference as a sum of infinitely many identical pieces.

The essence of this rebuttal is the assertion that even a very small nonzero number becomes arbitrarily large if it is added to itself enough times. Because the assertion was first made explicit by Archimedes, it is called the Archimedean property of the real numbers. An infinitesimal, if it existed, would be precisely a non-Archimedean number: a number greater than zero, which nevertheless remained less than 1, say, no matter how (finitely) many times it was added to itself. Archimedes, working in the tradition of Aristotle and Euclid, asserted that every number is Archimedean; there are no infinitesimals. Archimedes, however, was also a natural philosopher, an engineer and a physicist. He used infinitesimals and his physical intuition to solve problems in the geometry of parabolas. Then, since infinitesimals "do

METHOD OF EXHAUSTION is employed to prove indirectly that the area of a circle with radius 1 is half its circumference. In the proof on the opposite page a polygon is circumscribed on the circle (*top*), creating a number of triangles for which the areas can be calculated readily. By increasing the number of sides of the polygon, as in polygon *B* and polygon *N*, the triangles increase in number and become thinner, and the difference in area of circle and polygon becomes smaller. The difference will never be zero, however, for a polygon having any finite number of sides. Standard analysis avoids this difficulty by stating that, as the number of sides increases to infinity, the polygon's area approaches the circle's area as a limit. Nonstandard analysis avoids the concept of limit for a more suggestive explanation using a polygon with infinitely many sides, each side having infinitesimal length.

POLYGON *A*





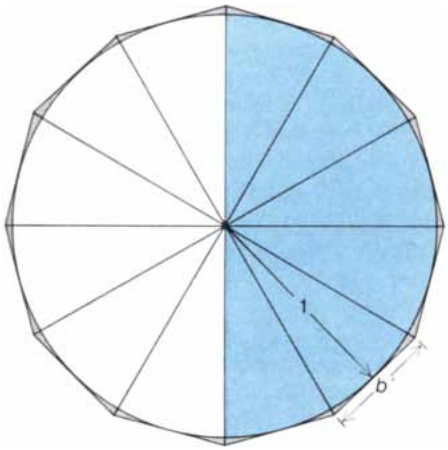
PERIMETER $P = 6b$

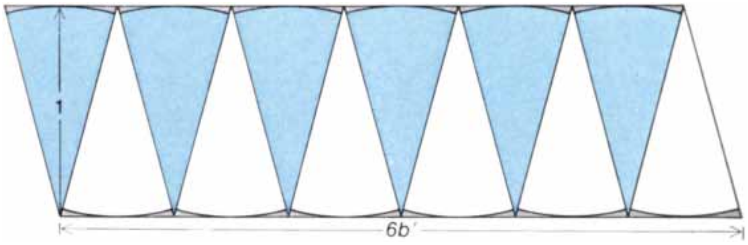
AREA OF POLYGON $A = 3b = \frac{1}{2}P$

CIRCUMFERENCE OF INSCRIBED CIRCLE $\ll P$

AREA OF INSCRIBED CIRCLE $\ll \frac{1}{2}P$

POLYGON *B*





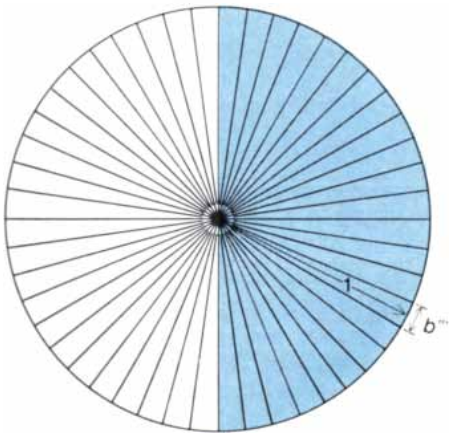
PERIMETER $P' = 12b'$

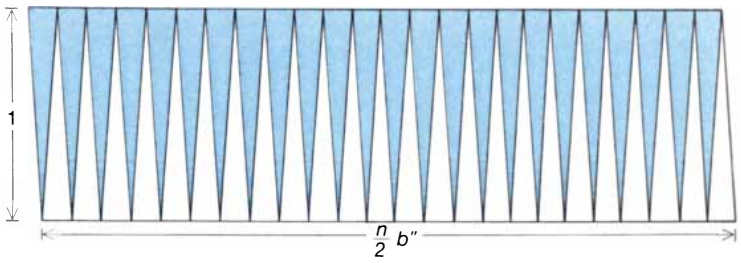
AREA OF POLYGON $B = 6b' = \frac{1}{2}P'$

CIRCUMFERENCE OF INSCRIBED CIRCLE $< P'$

AREA OF INSCRIBED CIRCLE $< \frac{1}{2}P$

POLYGON *N*





PERIMETER $P'' = nb''$

AREA OF POLYGON $N = \frac{nb''}{2} = \frac{1}{2}P''$

CIRCUMFERENCE OF INSCRIBED CIRCLE $\approx P''$

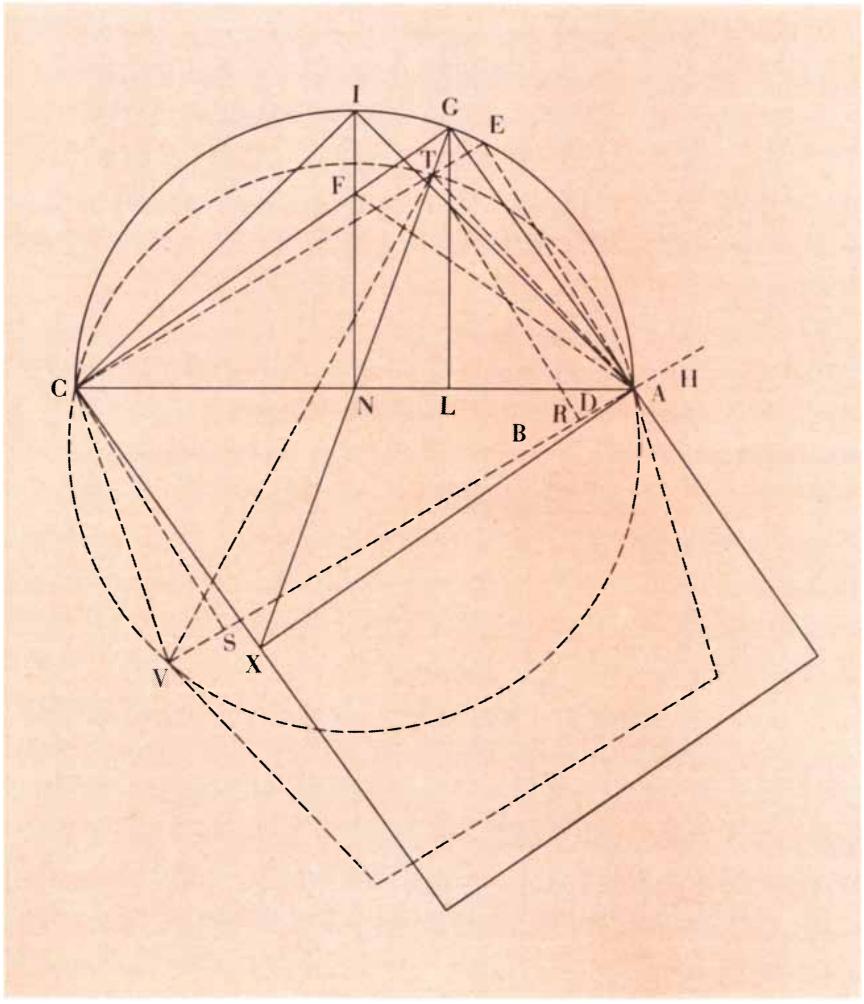
AREA OF INSCRIBED CIRCLE $\approx \frac{1}{2}P''$

not exist," he gave a "rigorous" proof of his results, using the "method of exhaustion," which relies on an indirect argument and purely finite constructions. The rigorous proof is given in his treatise *On the Quadrature of the Parabola,* which has been known since antiquity. The use of infinitesimals, which actually served to discover the answer, is in a paper called "On the Method," which was unknown until its sensational discovery in 1906.

Archimedes' method of exhaustion, which avoids infinitesimals, is in spirit close to the "epsilon-delta" method with which Weierstrass and his followers in the 19th century drove infinitesimal methods out of analysis. It is easy to explain if we refer to our example of the circle as an infinite-sided polygon. We wish to get a logically acceptable proof of the formula "The area of a circle with a radius of one unit equals half the circumference," which we discovered by a logically unacceptable argument.

We reason as follows. The formula asserts the equality of two quantities associated with a circle with a radius of 1: its area and half its circumference. Thus if the formula is false, one of these quantities is larger than the other. Let A be the positive number obtained by subtracting the smaller from the larger. Now, we can circumscribe about the circle a regular polygon with as many sides as we wish. Since the polygon is composed of a finite number of finite triangles with altitude 1, we know that its area is half its perimeter. By making the number of sides sufficiently large we can arrange for the polygon's area to differ from the area of the circle by less than half of A (whatever its value is taken to be); at the same time the perimeter of the polygon will differ from the perimeter of the circle by less than half of A. But then the area and the semiperimeter of the circle must differ by less than A, which contradicts the supposition from which we started. Hence the supposition is impossible and A must be zero, as we wished to prove.

This argument is logically impeccable. Compared with the directness of the first analysis, however, there is something fussy, even pedantic, about it. After all, if the use of infinitesimals gives the right answer, must not the argument be correct in some sense? Even if we cannot justify the concepts it employs, how can it really be wrong if it works?



PROBLEM OF WINE CASK was attacked by Johannes Kepler by means of infinitesimals in his *Nova stereometria doliorum vinariorum,* published in Austria in 1615 and 1616. The problem Kepler set for himself was finding the best proportions for a wine cask. A page from a 19th-century reproduction of the paper that was published in Europe is shown.

Such a defense of infinitesimals was not made by Archimedes. Indeed, in "On the Method" he is careful to explain that "the fact here stated is not actually demonstrated by the argument used" and that a rigorous proof had been published separately. On the other hand, Nicholas of Cusa, who was a cardinal of the church, preferred the reasoning by infinite quantities because of his belief that the infinite was "the source and means, and at the same time the unattainable goal, of all knowledge." Nicholas was followed in his mysticism by Johannes Kepler, one of the founders of modern science. In a work less well known nowadays than his discoveries in astronomy, Kepler in 1612 used infinitesimals to find the best proportions for a wine cask. He was not troubled by the self-contradictions in his method; he relied on divine inspiration, and he wrote that "nature teaches geometry by instinct alone, even without ratiocination." Moreover, his formulas for the volumes of wine casks are correct.

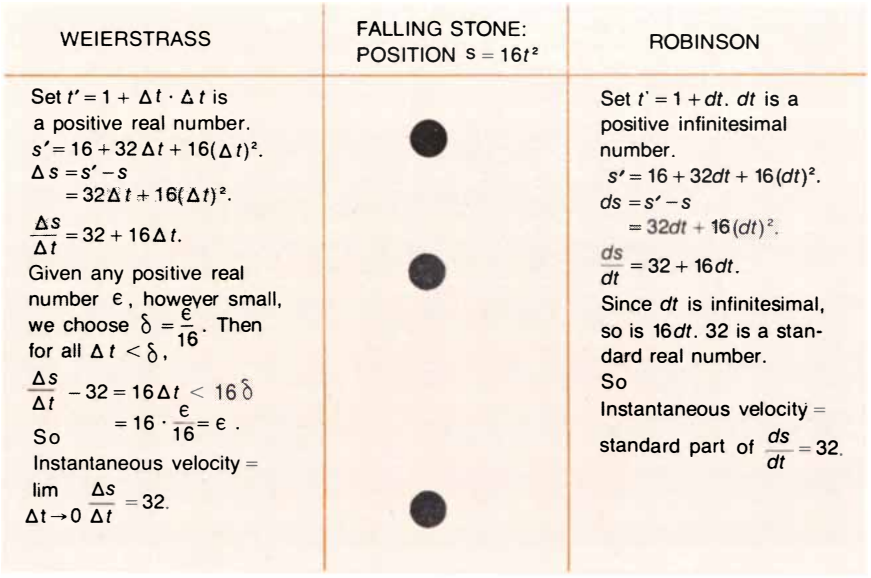
The most famous mathematical mystic was no doubt Blaise Pascal. In answering those of his contemporaries who objected to reasoning with infinitely small quantities, Pascal was fond of saying that the heart intervenes to make the work clear. Pascal looked on the infinitely large and the infinitely small as mysteries, something that nature has proposed to man not for him to understand but for him to admire.

The full flower of infinitesimal reasoning came with the generations after Pascal: Newton, Leibniz, the Bernoulli brothers (Jakob and Johann) and Leonhard Euler. The fundamental theorems of the calculus were found by Newton and Leibniz in the 1660's and 1670's. The first textbook on the calculus was written in 1696 by the Marquis de L'Hôpital, a pupil of Leibniz' and Johann Bernoulli's. Here it is stated at the

outset as an axiom that two quantities differing by an infinitesimal can be considered to be equal. In other words, the quantities are at the same time considered to be equal to each other and not equal to each other! A second axiom states that a curve is "the totality of an infinity of straight segments, each infinitely small." This is an open embracing of methods that Aristotle had outlawed 2,000 years earlier.

Indeed, wrote L'Hôpital, "ordinary analysis deals only with finite quantities; this one penetrates as far as infinity itself. It compares the infinitely small differences of finite quantities; it discovers the relations between these differences, and in this way makes known the relations between finite quantities that are, as it were, infinite compared with the infinitely small quantities. One may even say that this analysis extends beyond infinity, for it does not confine itself to the infinitely small differences but discovers the relations between the differences of these differences."

Newton and Leibniz did not share L'Hôpital's enthusiasm. Leibniz did not claim that infinitesimals really existed, only that one could reason without error as if they did exist. Although Leibniz could not substantiate this claim, Robinson's work shows that in some sense he was right after all. Newton tried to avoid the infinitesimal. In his *Principia Mathematica,* as in Archimedes' *On the Quadrature of the Parabola,* results that were originally found by infinitesimal methods are presented in a purely finite Euclidean fashion.



FALLING-STONE PROBLEM is depicted as it would be solved by standard analysis (*left*) and nonstandard analysis (*right*). Standard analysis, exemplified by the 19th-century German mathematician Karl Weierstrass, computes the velocity of the falling stone at any instant without employing infinitesimals, defining the speed instead as a limit that is approximated by ratios of finite increments. Abraham Robinson of Yale University, who invented nonstandard analysis, makes the computation with a modified infinitesimal method.

Dynamics had become as important as geometry in providing questions for mathematical analysis. The leading problem was the connection between "fluents" and "fluxions," what would today be called the instantaneous position and the instantaneous velocity of a moving body.

Consider a falling stone. Its motion is described by giving its position as a function of time. As it falls its velocity increases, so that the velocity at each instant is also a variable function of time. Newton called the position function the "fluent" and the velocity function the "fluxion." If either of the two is given, the other can be determined; this connection is the heart of the infinitesimal calculus fashioned by Newton and Leibniz.

In the case of a falling stone the fluent is given by the formula $s = 16t^2$, where s is the number of feet traveled and t is the number of seconds elapsed since the stone was released. As the stone falls its velocity increases steadily. How can we compute the velocity of the falling stone at some instant of time, say at $t = 1$?

We could find the *average* velocity for a finite time by the elementary formula: velocity equals distance divided by time. Can we use this formula to find the instantaneous velocity? In an infinitesimal increment of time the increment of distance would also be infinitesimal; their ratio, the average speed during the instant, should be the finite instantaneous velocity we seek.

We let dt stand for the infinitesimal increment of time and ds for the corresponding increment of distance. (Of course ds and dt must be thought of as single symbols and not as d times t or d times s.) We want to find the ratio ds/dt, which is to be finite. To find the increment of distance from $t = 1$ to $t = 1 + dt$ we compute the position of the stone when $t = 1$, which is $16 \times 1^2 = 16$, and its position when $t = 1 + dt$, which is $16 \times (1 + dt)^2$. Using a little elementary algebra, we find that ds, the increment of distance, which is the difference of these two distances, is $32dt + 16dt^2$. Thus the ratio ds/dt, which is the quantity we are trying to find, is equal to $32 + 16dt$.

Have we solved our problem? Since the answer should be a finite quantity, we should like to drop the infinitesimal term, $16dt$, and get the answer, 32 feet per second, for the instantaneous velocity. That is precisely what Bishop Berkeley will not let us do.

The Analyst, Berkeley's brilliant and devastating critique of the infinitesimal method, appeared in 1734. The book was addressed to "an infidel mathematician," who is generally supposed to have been Newton's friend the astronomer Edmund Halley. Halley financed the publication of the *Principia* and helped to prepare it for the press. It is said that he also persuaded a friend of Berkeley's of the "inconceivability of the doctrines of Christianity"; the Bishop responded that Newton's fluxions were as "obscure, repugnant and precarious" as any point in divinity.

"I shall claim the privilege of a Freethinker," wrote the Bishop, "and take the liberty to inquire into the object, principles, and method of demonstration admitted by the mathematicians of the present date, with the same freedom that you presume to treat the principles and mysteries of Religion." Berkeley declared that the Leibniz procedure, simply "considering" $32 + 16dt$ to be "the same" as 32, was unintelligible. "Nor will it avail," he wrote, "to say that [the term neglected] is a quantity exceedingly small; since we are told that *in rebus mathematicis errores quam minimi non sunt contemnendi.*" If something is neglected, no matter how small, we can no

longer claim to have the exact velocity but only an approximation.

Newton, unlike Leibniz, tried in his later writings to soften the "harshness" of the doctrine of infinitesimals by using physically suggestive language. "By the ultimate velocity is meant that with which the body is moved, neither before it arrives at its last place, when the motion ceases, nor after; but at the very instant when it arrives.... And, in like manner, by the ultimate ratio of evanescent quantities is to be understood the ratio of the quantities, not before they vanish, nor after, but that with which they vanish." When he proceeded to compute, however, he still had to justify dropping unwanted "negligible" terms from his computed answer. Newton's argument was to find first, as we have done, $ds/dt = 32 + 16dt$, and then to set the increment dt equal to zero, leaving 32 as the exact answer.

But, wrote Berkeley, "it should seem that this reasoning is not fair or conclusive." After all, dt is either equal to zero or not equal to zero. If dt is not zero, then $32 + 16dt$ is not the same as 32. If dt is zero, then the increment in distance ds is also zero, and the fraction ds/dt is not $32 + 16dt$ but a meaningless expression, 0/0. "For when it is said, let the increments vanish, i.e., let the increments be nothing, or let there be no increments, the former supposition that the increments were something, or that there were increments, is destroyed, and yet a consequence of that supposition, i.e., an expression got by virtue thereof, is retained. Which is a false way of reasoning." Berkeley charitably concluded: "What are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?"

Berkeley's logic could not be answered; nevertheless, mathematicians went on using infinitesimals for another century, and with great success. Indeed, physicists and engineers have never stopped using them. In pure mathematics, on the other hand, a return to Euclidean rigor was achieved in the 19th century, culminating under the leadership of Weierstrass in 1872. It is interesting to note that the 18th century, the great age of the infinitesimal, was the time when no barrier between mathematics and physics was recognized. The leading physicists and the leading mathematicians were the same people. When pure mathematics reappeared as a separate discipline, mathematicians again made sure that the foundations of their work contained no obvious contradictions. Modern analysis secured its foundations by doing what the Greeks had done: outlawing infinitesimals.

To find an instantaneous velocity according to the Weierstrass method we abandon any attempt to compute the speed as a ratio. Instead we define the speed as a limit, which is approximated by ratios of finite increments. Let Δt be a variable finite time increment and Δs be the corresponding variable space increment. Then $\Delta s/\Delta t$ is the variable quantity $32 + 16\Delta t$. By choosing Δt sufficiently small we can make $\Delta s/\Delta t$ take on values as close as we like to the value 32, and so, by definition, the speed at $t = 1$ is exactly 32.

This approach succeeds in removing any reference to numbers that are not finite. It also avoids any attempt directly to set Δt equal to zero in the fraction $\Delta s/\Delta t$. Thus we avoid both of the logical pitfalls exposed by Bishop Berkeley. We do, however, pay a price. The intuitively clear and physically measurable quantity, the instantaneous velocity, becomes subject to the surprisingly subtle notion of "limit." If we spell out in detail what that means, we have the following tongue-twister:

"The velocity is v if, for any positive number ε, $\Delta s/\Delta t - v$ is less than ε in absolute value for all values of Δt less in absolute value than some other positive number δ (which will depend on ε and t)."

We have defined v by means of a subtle relation between two new quantities, ε and δ, which in some sense are irrelevant to v itself. At least ignorance of ε and δ never prevented Bernoulli or Euler from finding a velocity. The truth is that in a real sense we already knew what instantaneous velocity was before we learned this definition; for the sake of logical consistency we accept a definition that is much harder to understand than the concept being defined. Of course, to a trained mathematician the epsilon-delta definition is intuitive; this shows what can be accomplished by proper training.

The reconstruction of the calculus on the basis of the limit concept and its epsilon-delta definition amounted to a reduction of the calculus to the arithmetic of real numbers. The momentum gathered by these foundational clarifications led naturally to an assault on the logical foundations of the real-number system itself. This was a return after two and a half millenniums to the problem of irrational numbers, which the Greeks had abandoned as hopeless after Pythagoras. One of the tools in these efforts was the newly developing field of mathematical, or symbolic, logic.

More recently it has been found that mathematical logic provides a conceptual foundation for the theory of computing machines and computer programs. Hence this prototype of purity in mathematics now has to be regarded as belonging to the applicable part of mathematics.

The link between logic and computing is to a great extent the notion of a formal language, which is the kind of language machines understand. And it is the notion of a formal language that enabled Robinson to make precise Leibniz' claim that one could without error reason as if infinitesimals existed.

Leibniz had thought of infinitesimals

SYMBOL	INTENDED MEANING
~	not
&	and
V	or
→	implies
∀	for all
∃	there exists
=	equals
x y z	variables ranging over real numbers
f g h	variables ranging over other objects
+ · <	plus, times, less than
( ) [ ]	parentheses
0 1 2	symbols for particular numbers

SYMBOLS EMPLOYED in the formal language L, in which calculus can be expressed, are translated into English in this partial dictionary. The formal language, which employs many more symbols than these, provides a link between the standard universe and the larger nonstandard universe that is a central concept of nonstandard mathematical analysis.

as being infinitely small positive or negative numbers that still had "the same properties" as the ordinary numbers of mathematics. On its face the idea seems self-contradictory. If infinitesimals have the same "properties" as ordinary numbers, how can they have the "property" of being positive yet smaller than any ordinary positive number? It was by using a formal language that Robinson was able to resolve the paradox. Robinson showed how to construct a system containing infinitesimals that was identical with the system of "real" numbers with respect to all those properties expressible in a certain formal language. Naturally the "property" of being positive yet smaller than any ordinary positive number will turn out *not* to be expressible in the language, thereby escaping the paradox.

The situation is familiar to anyone who has ever communicated with a computing machine. A computer accepts as inputs only symbols from a certain list that is given in advance to the user, and the symbols must be used in accordance with certain given rules. Ordinary language, as used in human communication, is subject to rules that linguists are still far from understanding. Computers are "stupid," if you have to communicate with them, precisely because unlike humans they work in a formal language with a given vocabulary and a given set of rules. Humans work in a natural language, with rules that have never been made fully explicit.

Mathematics, of course, is a human activity, like philosophy or the design of computers; like these other activities, it is carried on by humans using natural languages. At the same time mathematics has, as a special feature, the ability to be well described by a formal language, which in some sense mirrors its content precisely. It might be said that the possibility of putting a mathematical discovery into a formal language is the test of whether it is fully understood.

In nonstandard analysis one takes as the starting point the finite real numbers and the rest of the calculus as known to standard mathematicians. Call this the "standard universe," designated by the letter M. The formal language in which we talk about M can be designated L. Any sentence in L is a proposition about M, and of course it must be either true or false. That is, any sentence in L is either true or its negation is true. We call the set of all true sentences K, and we say M is a "model" for K. By this we mean that M is a mathematical structure such that every sentence in K, when interpreted as referring to M, is true. Of course, we do not "know" K in any effective sense; if we did, we would have the answer to every possible question in analysis. Nevertheless, we regard K as being a well-defined object, about which we can reason and draw conclusions.

FORMAL SENTENCE IN L	INTERPRETATION IN STANDARD UNIVERSE	INTERPRETATION IN NONSTANDARD UNIVERSE
$(\forall x)(\exists y)[x = 0 \vee xy = 1]$ Literally: For all x, there exists y such that either $x = 0$ or $xy = 1$.	Every nonzero real number has a reciprocal.	Every nonzero nonstandard real number has a nonstandard reciprocal; in particular positive infinitesimals have reciprocals that are larger than any standard, real number, i.e., they are infinite.

FORMAL SENTENCE is stated in the language L. The middle column gives its interpretation or meaning in the standard universe; right-hand column, in the nonstandard universe.

The essential fact, the main point, is that in addition to M, the standard universe, there are also nonstandard models for K. That is, there are mathematical structures M^*, essentially different from M (in a sense we shall explain) and that nevertheless are models for K in the natural sense of the term: there are objects in M^* and relations between objects in M^* such that if the symbols in L are reinterpreted to apply to these pseudo-objects and pseudo-relations in the appropriate way, then every sentence in K is still true, although with a different meaning.

A crude analogy may help the intuition. Let M be the set of graduating seniors at Central High School. Suppose, for argument's sake, that all these students had their picture taken for the yearbook, where the students all appear in two-inch squares. Then M^* can be the set of all two-inch squares on any page of the yearbook. Clearly, with an obvious interpretation, any true statement about a student at Central High corresponds to a true statement about a certain two-inch square in the yearbook. Still, there are many two-inch squares in the yearbook that do not correspond to any student. M^* is much bigger than M; in addition to members corresponding to the members of M, it also contains many other members.

Hence the statement "Harry Smith is thinner than George Klein," when interpreted in M^*, is a statement about certain two-inch squares. It is not true if the relation "thinner than" is interpreted in the standard way. Thus "thinner than" has to be reinterpreted, as a pseudo-relation, between pseudo-students (pictures of students). We could define the pseudo-relation "thinner than" (in quotation marks) by saying that the two-inch square labeled "Harry Smith" is "thinner than" the two-inch square labeled "George Klein" only if Harry Smith is actually thinner than George Klein. In this way true statements about students are reinterpreted as true statements about two-inch squares.

Of course, in this example the entire argument is a bit contrived. If M is the standard universe for the calculus, however, then M^*, the nonstandard universe, is a remarkable and interesting place.

The existence of interesting nonstandard models was first discovered by the Norwegian logician Thoralf A. Skolem, who found that the axioms of counting—the axioms that describe the "natural numbers" 1, 2, 3 and so on—have nonstandard models containing "strange" objects not contemplated in ordinary arithmetic. Robinson's great insight was to see how this exotic offshoot of modern formal logic could be the basis for resurrecting infinitesimal methods in differential and integral calculus. In this resurrection he relied on a theorem first proved by the Russian logician Anatoli Malcev and then generalized by Leon A. Henkin of the University of California at Berkeley. This is the "compactness" theorem. It is related to the famous "completeness" theorem of Kurt Gödel, which states that a set of sentences is logically consistent (no contradiction can be deduced from the sentences) if and only if the sentences have a model, that is, if and only if there is a "universe" in which they are all true.

The compactness theorem states the following: Suppose we have a collection of sentences in the language L. Suppose in the standard universe every finite subset of this collection is true. Then there exists a nonstandard universe where the entire collection is true at once.

The compactness theorem follows easily from the completeness theorem: if

every finite subset of a collection of sentences of L is true in the standard universe, then every finite subset is logically consistent. So the entire collection of sentences is logically consistent (since any deduction can make use of only a finite number of premises). By the completeness theorem there is a (nonstandard) universe in which the entire collection is true.

A direct consequence of the compactness theorem is the "existence" of infinitesimals. To see how this amazing result follows from the compactness theorem consider the sentences:

"C is a number bigger than zero and less than $1/2$."

"C is a number bigger than zero and less than $1/3$."

"C is a number bigger than zero and less than $1/4$." And so on.

This is an infinite collection of sentences each of which can be written in the formal language L. With reference to the standard universe R of real numbers, every finite subset is true, because if you have finitely many sentences of the form "C is a number bigger than zero and less than $1/n$," then one of the sentences will contain the smallest fraction $1/n$, and $1/2n$ will indeed be bigger than zero and smaller than all the fractions in your finite list of sentences. And yet if you consider the entire infinite set of these sentences, it is false with reference to the standard real numbers, because no matter how small a positive real number c you choose, $1/n$ will be smaller than c if n is big enough.

The compactness theorem of Malcev and Henkin states that there is a nonstandard universe containing pseudo-reals R^* including a positive pseudo-real number c smaller than any number of the form $1/n$. That is, c is infinitesimal. Moreover, c has all the properties of standard real numbers in a perfectly precise sense: any true statement about the standard reals that you can state in the formal language L is true also about the nonstandard reals, including the infinitesimal c—under the appropriate interpretation. (The two-inch square labeled "Harry Smith" is not really thinner than the two-inch square labeled "George Klein," but the statement "Harry Smith" is "thinner than" "George Klein" is true, under our nonstandard interpretation of "thinner than." On the other hand, properties shared by all the standard real numbers may not apply to the nonstandard pseudo-numbers, if these properties cannot be expressed in the formal language L.

The Archimedean property (nonexistence of infinitesimals) of R can be expressed by using an infinite set of sentences of L as follows (we use the symbol ">" as usual to mean "is greater than"). For each positive element c of R all but a finite number of the sentences below are true:

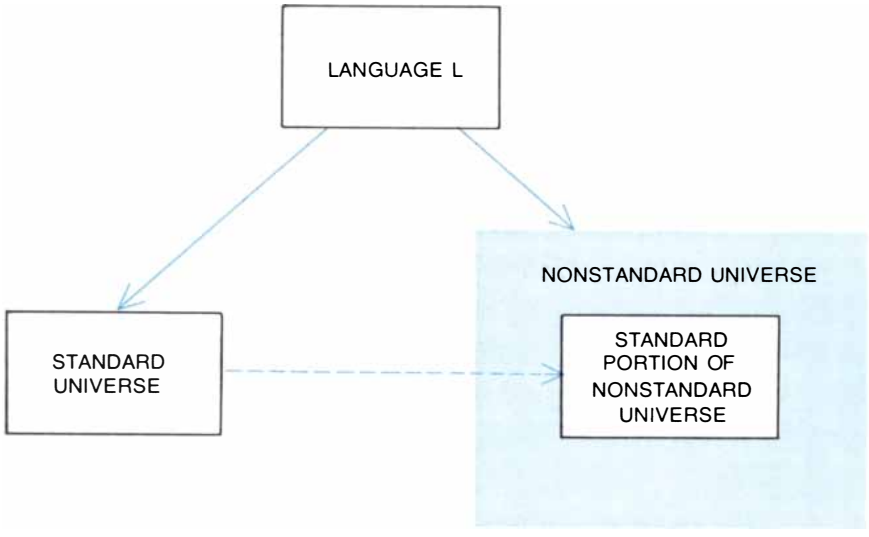
$$c > 1$$
$$c + c > 1$$
$$c + c + c > 1, \text{ and so on.}$$

This is not true, however, for the pseudo-reals R^*: if c is infinitesimal (hence pseudo-real), all these sentences are false. In other words, no finite sum of c's can exceed 1, no matter how many terms we take. The very fact that the Archimedean property is true in the standard world but false in the nonstandard one proves that the property cannot be expressed by a sentence of L; the statement we have used involves infinitely many sentences. It is precisely this distinction that makes the pseudo-objects useful. They behave "formally" like standard objects and yet they differ with respect to important properties that are not formalized by L.

Although the nonstandard universe is conceptually distinct from the standard one, it is desirable to think of it as an enlargement of the standard universe. Since R^* is a model for L, every true sentence about R has an interpretation in R^*. In particular the names of numbers in R have an interpretation as names of objects in R^*. We can simply identify the object in R^* called "2" with the familiar number 2 in R. Then R^* contains the standard real numbers in R, along with a vast collection of infinitesimal and infinite quantities, in which R is embedded.

An object in R^* (a pseudo-real number) is called infinite if it is pseudo-greater than every standard real number; otherwise it is called finite. A positive pseudo-real number is called infinitesimal if it is pseudo-smaller than every positive standard real number. If the pseudo-difference of two pseudo-reals is finite, we say they belong to the same "galaxy"; the pseudo-real axis contains an uncountable infinity of galaxies. If the pseudo-difference of two pseudo-reals is infinitesimal, we say they belong to the same "monad" (a term Robinson borrowed from Leibniz' philosophical writings). If a pseudo-real r^* is infinitely close to a standard real number r, we say r is the standard part of r^*. All the standard reals are of course in the same galaxy, which is called the principal galaxy. In the principal galaxy every monad contains one and only one standard real number. This monad is the "infinitesimal neighborhood" of r: the set of nonstandard reals infinitely close to r. The notion of a monad turns out to be applicable not only to real numbers but also to general metric and topological spaces. Nonstandard analysis therefore is relevant not just to elementary calculus but to the entire range of modern abstract analysis.

When we say infinitesimals or monads exist, it should be clear that we do not



ROLE OF FORMAL LANGUAGE in mediating between standard and nonstandard universes is portrayed. Formal language L describes the standard universe, which includes the real numbers of classical mathematics. Sentences of L that are true in standard universe are also true in nonstandard one, which contains additional mathematical objects such as infinitesimals. Nonstandard analysis thus makes the infinitesimal method precise for first time.

The Chevy we rent you in San Francisco is just as nice.

We like to baby our cars.

Because back when we were just a small car rental company we realized something. That people enjoy driving a really sharp-looking car.

So we gave them sharp-looking cars. And they gave us their business.

They gave us so much business that we're now renting more new GM cars than anyone else in the world.

But other than that, we haven't changed.

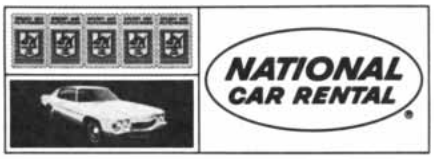
We're still faithful to your reservation as though it were sacred.

We still thank you for your business with more than just a thank you. We give you good old S&H Green Stamps.*

And as always, no matter where in the world you rent one of our cars we're determined to rent you only one kind.

The kind you'd be happy to park in your own driveway.

To reserve a car in the U.S. and all over the world call 800-328-4567 toll free.†



Take advantage of us.

National Car Rental System (in Canada it's Tilden Rent-A-Car). *Valid on U.S. rentals only except where prohibited by law. †In Minnesota call 612-944-1234 collect. © National Car Rental System, Inc. 1972

mean this at all in the sense it would have been understood by Euclid or by Berkeley. Until 100 years ago it was tacitly assumed by all philosophers and mathematicians that the subject matter of mathematics was objectively real in a sense close to the sense in which the subject matter of physics is real. Whether infinitesimals did or did not exist was a question of fact, not too different from the question of whether material atoms do or do not exist. Today many, perhaps most, mathematicians have no such conviction of the objective existence of the objects they study. Model theory entails no commitment one way or the other on such ontological questions. What mathematicians want from infinitesimals is not material existence but rather the right to use them in proofs. For this all one needs is the assurance that a proof using infinitesimals is no worse than one free of infinitesimals.

The employment of nonstandard analysis in research goes something like this. One wishes to prove a theorem involving only standard objects. If one embeds the standard objects in the nonstandard enlargement, one may be able to find a much shorter and more "insightful" proof by using nonstandard objects. The theorem has then been proved actually with reference to the nonstandard interpretation of its words and symbols. Those nonstandard objects that correspond to standard objects have the feature that sentences about them are true (in the nonstandard interpretation) only if the same sentence is true with reference to the standard object (in the standard interpretation). Thus we prove theorems about standard objects by reasoning about nonstandard objects.

For example, recall Nicholas of Cusa's "proof" that the area of a circle with a radius of 1 equals half its circumference. In Robinson's theory we see in what sense Nicholas' argument is correct. Once infinitesimal and infinite numbers are available (in the nonstandard universe) it can be proved that the area of the circle is the standard part of the sum (in the nonstandard universe) of infinitely many infinitesimals.

Here is how the falling-stone problem would look according to Robinson. We define the instantaneous velocity not as the ratio of infinitesimal increments, as L'Hôpital did, but rather as the standard part of that ratio; then ds, dt and their ratio ds/dt are nonstandard real numbers. We have as before $ds/dt = 32 + 16dt$, but now we immediately conclude, rigorously and without any limiting argument, that v, the standard part of ds/dt, equals 32. A slight modification in the Leibniz method of infinitesimals, distinguishing carefully between the nonstandard number ds/dt and its standard part v, avoids the contradiction, which L'Hôpital simply ignored.

Of course, a proof is required that the Robinson definition gives the same answer in general as the Weierstrass definition. The proof is not difficult, but we shall not attempt to give it here.

What is achieved is that the infinitesimal method is for the first time made precise. In the past mathematicians had to make a choice. If they used infinitesimals, they had to rely on experience and intuition to reason correctly. "Just go on," Jean Le Rond d'Alembert is supposed to have assured a hesitating mathematical friend, "and faith will soon return." For rigorous certainty one had to resort to the cumbersome Archimedean method of exhaustion or its modern version, the Weierstrass epsilon-delta method. Now the method of infinitesimals, or more generally the method of monads, is elevated from the heuristic to the rigorous level. The approach of formal logic succeeds by totally evading the question that excited Berkeley and all the other controversialists of former times, that is, whether or not infinitesimal quantities really exist in some objective sense.

From the viewpoint of the working mathematician the important thing is that he regains certain methods of proof, certain lines of reasoning, that have been fruitful since before Archimedes. The notion of an infinitesimal neighborhood is no longer a self-contradictory figure of speech but a precisely defined concept, as legitimate as any other in analysis.

The applications we have discussed are elementary, in fact trivial. Nontrivial applications have been and are being made. Work has appeared on nonstandard dynamics and nonstandard probability. Robinson and his pupil Allen Bernstein used nonstandard analysis to solve a previously unsolved problem on compact linear operators. It must nonetheless be said that many analysts remain skeptical about the ultimate importance of Robinson's method. It is quite true that whatever can be done with infinitesimals can in principle be done without them. Perhaps, as with other radical innovations, the full use of the new ideas will be made by a new generation of mathematicians who are not too deeply embedded in standard methods to enjoy the freedom and power of nonstandard analysis.

Six facts you won't read in an ad for any other steel belted radial tire.

1 THE PIRELLI CINTURATO CN-75 IS THE ONLY TIRE TO USE "TRAC-STEEL 21™". STEEL CORDS COMPRISED OF A VIRTUALLY UNHEARD OF NUMBER OF STEEL FILAMENTS.

In the Cinturato steel belted radial, each steel cord is made up of 21 individual filaments.

We know of no other tire that uses that many. Although they should.

Because the more filaments of steel used in a cord, the greater the directional flexibility of the tire.

Or to put it in layman's terms, you get better traction, better handling and a more comfortable ride.

2 THE CN-75 IS THE ONLY TIRE TO COMBINE THE STRENGTH AND PROTECTION OF STEEL WITH THE EFFORTLESS HANDLING AND QUIET COMFORT OF TEXTILE.

Because of its unique combination of "Trac-Steel 21," and rayon, the Pirelli Cinturato has handling and comfort characteristics not usually associated with steel belted tires.

This unique combination enables the Cinturato to corner, maneuver and grip the road like the treads of a tank.

Yet ride smoothly and silently. Without harshness. Without hum. Without vibration. On any surface. At any speed.

3 THE CN-75 IS THE ONLY TIRE THAT WAS NOT ONLY DESIGNED FOR AMERICAN CARS, IT WAS "TUNED" TO THEM.

Like many other steel belted radials, the Cinturato was designed for American car suspensions.

But then we at Pirelli put the Cinturato through a performance test. For over a million miles. Under all kinds of driving conditions.

We tested our tire on American roads. On virtually every class of American car. From the mini cars to the luxury cars to the high performance jobs.

Every aspect of the tire was tested. And if anything needed improving along the way, we improved it.

Before we introduced it to you.

4 BEFORE THE CN-75 WAS INTRODUCED, WE ROAD TESTED IT FOR OVER A MILLION MILES. TO BE SURE WE COULD GUARANTEE YOU AT LEAST 40,000.

Now don't think that every car we tested got 40,000 miles. Some only got 30,000. Some got over 50,000. (When you test under conditions as tough as we did there are bound to be variations.)

But we are sure we can guarantee you at least 40,000 miles.* And there's something else different about Pirelli's guarantee.

While we strongly recommend that you check your front end, and do a few other things every 6,000 miles or so (for your own safety), we won't cancel your guarantee if you don't. As long as any problem is unrelated to a vehicle defect.

In other words, we didn't just guarantee our tire and hope you wouldn't read the fine print.

5 EVERY SINGLE CN-75 IS COMPLETELY X-RAY INSPECTED. AFFORDING US 100% QUALITY CONTROL.

We at Pirelli have been making radials for over 20 years, and we know that a steel belted radial is very difficult to make.

So we don't take any chances.

Each CN-75 is X-rayed. 100%.

Because we know that even if the slightest thing is wrong with the tire, the effect it could have on a car will be more than slight.

6 THE PIRELLI CINTURATO CN-75 NEVER HAS TO BE ROTATED.

If you keep your car in reasonably good shape, tires properly inflated and wheels balanced, Pirelli Cinturatos will wear evenly.

The reason is that Cinturatos are so precisely built that it makes uneven wear virtually impossible.

So there's no need to rotate them.

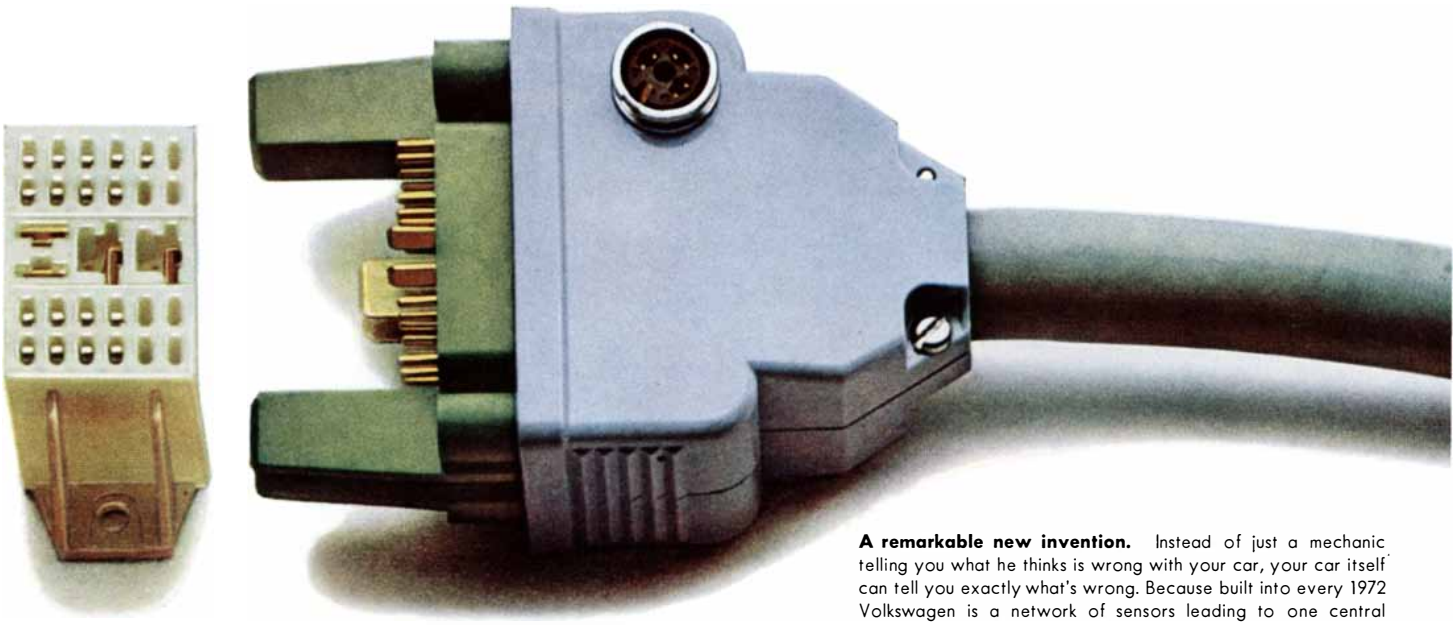
See if any tire company will tell you *that* about their tire.

Or, for that matter, any of the other things we've mentioned.

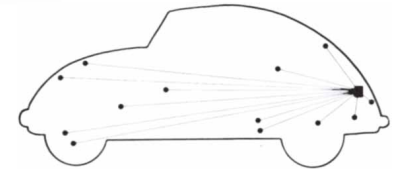
PIRELLI
Steel Belted Radial Tires

*PIRELLI TIRE CORPORATION'S GUARANTEE FOR THE CN75 RADIAL HIGHWAY TUBELESS WHITEWALL TIRE COVERS [illegible] DEFECTS IN MATERIALS AND WORKMANSHIP AND NORMAL ROAD HAZARDS (EXCLUDING REPAIRABLE PUNCTURES) WHEN TIRE IS USED ON DOMESTIC (AMERICAN MANUFACTURED) PASSENGER VEHICLES IN NORMAL SERVICE WITHIN THE CONTINENTAL UNITED STATES (EXCLUDING ALASKA). UNDER THE TERMS OF THE MILEAGE GUARANTEE PIRELLI TIRE CORPORATION WILL GIVE CREDIT OR REFUND (AT PIRELLI'S OPTION) EQUAL TO THE CURRENT NATIONAL PRICE MULTIPLIED BY THE PERCENTAGE OF GUARANTEED MILEAGE NOT RUN ON THE TIRE. GUARANTEE EFFECTIVE MAY 15, 1972. PIRELLI TIRE CORPORATION, SIXTY EAST FORTY-SECOND ST. NEW YORK, N.Y. 10017

Try looking at a Volkswagen this way:



A remarkable new invention. Instead of just a mechanic telling you what he thinks is wrong with your car, your car itself can tell you exactly what's wrong. Because built into every 1972 Volkswagen is a network of sensors leading to one central socket which when plugged into a computer (that VW Dealers will have starting later this year) can report the condition of vital parts in your VW directly to you—all printed out in plain English on a sheet for you to keep.



An old reason to buy a new one. We build VWs to last. And so far, the record belongs to Jim Dillon of Bluefield, West Va. His '55 VW (with the help of 2 rebuilt engines) has gone well over the distance of from here to the moon and back.

Who lost the least?†	Depreciation as of Jan., 1972.
 1969 Nova 4 Sedan 2 Dr.	–$814
 1969 Opel 2 Dr. Sedan	–$812
 1969 Datsun PL 510 2 Dr.	–$736
 1969 Toyota Corolla Sedan 2 Dr.	–$686
 1969 Volkswagen 113	–$449

Worth more when you sell it.† Consider this advanced thinking. The real price of a car is the cash difference between what you buy it for new and what you sell it for used. And based on the past, after 3 or 4 years, no other economy car gives you back more money than a VW.

After we paint the car, we paint the paint. We cover our car with 13 pounds of paint—inside and out. Including places you can't see, but which corrosion can find. We don't expect the paint to hold the VW together. But it will help keep it from falling apart.

*If an owner maintains and services his vehicle in accordance with the Volkswagen Maintenance Schedule, any factory part found to be defective in material or workmanship within 24 months or 24,000 miles, whichever comes first (except normal wear and tear and service items), will be repaired or replaced by any U.S. or Canadian Volkswagen Dealer. And this will be done free of charge. See your dealer for details.

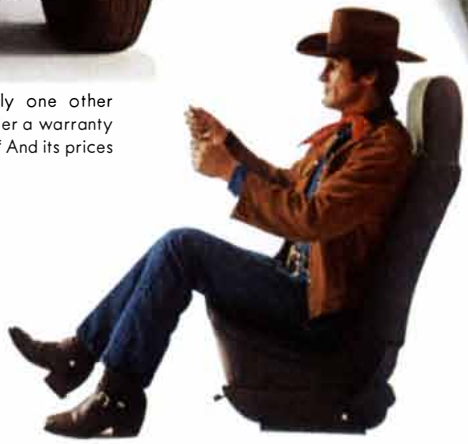
Inspected by a cast of thousands. Before it can leave our factory, every new VW has to face 1104 inspectors. One "no" from any inspector and it doesn't go—until it's right.

After we seal the top, we seal the bottom. The VW is so advanced, it has a sealed steel bottom. To protect its insides against everything outside.

The only car in the world with a longer warranty. Only one other car besides the VW considers itself advanced enough to offer a warranty that exceeds today's standard 12-month or12,000-mile limits.* And its prices start at $25,200 while our price is only $1,999.**

The 1972 Volkswagen Beetle is more than a simple, honest, dependable car.

It's a simple, honest, dependable car that's years ahead of its time.



More room than you think. As small as our car is outside, inside it has more than enough headroom and legroom for the average-size man. Even when he's wearing a not-so-average-size hat.

**1972 Volkswagen Sedan 111 suggested retail price, P.O.E., local taxes and other dealer charges, if any, additional. ©Volkswagen of America, Inc.
†Based on 1969 manufacturers' suggested retail prices and 1972 average resale prices as quoted in NADA Official Used Car Guide, Eastern Edition, Jan. 1972

CONTOUR AND CONTRAST

We see contours when adjacent areas contrast sharply. Surprisingly, certain contours, in turn, make large areas appear lighter or darker than they really are. What neural mechanisms underlie these effects?

by Floyd Ratliff

Contours are so dominant in our visual perception that when we draw an object, it is almost instinctive for us to begin by sketching its outlines. The use of a line to depict a contour may well have been one of the earliest developments in art, as exemplified by the "line drawings" in the pictographs and petroglyphs of prehistoric artists. We see contours when there is a contrast, or difference, in the brightness or color between adjacent areas. How contrast creates contours has been thoroughly studied by both scientists and artists. How the contour itself can affect the contrast of the areas it separates has been known to artists for at least 1,000 years, but it is relatively new as a subject of scientific investigation. Although the psychophysiological basis of how contrast enables the visual system to distinguish contours has been studied for the past century, it is only in the past few years that psychologists and physiologists have started to examine systematically the influence of contour on contrast.

You can readily observe how the visual system tends to abstract and accentuate contours in patterns of varying contrast by paying close attention to the edges of a shadow cast by an object in strong sunlight. Stand with your back to the sun and look closely at the shadow of your head and shoulders on a sidewalk. You will see a narrow half-shadow between the full shadow and the full sunlight. Objectively the illumination in the full shadow is uniformly low, in the half-shadow it is more or less uniformly graded and in the full sunlight it is uniformly high; within each area there are no sharp maxima or minima. Yet you will see a narrow dark band at the dark edge of the half-shadow and a narrow bright band at its bright edge. You can enhance the effect by swaying from side to side to produce a moving shadow.

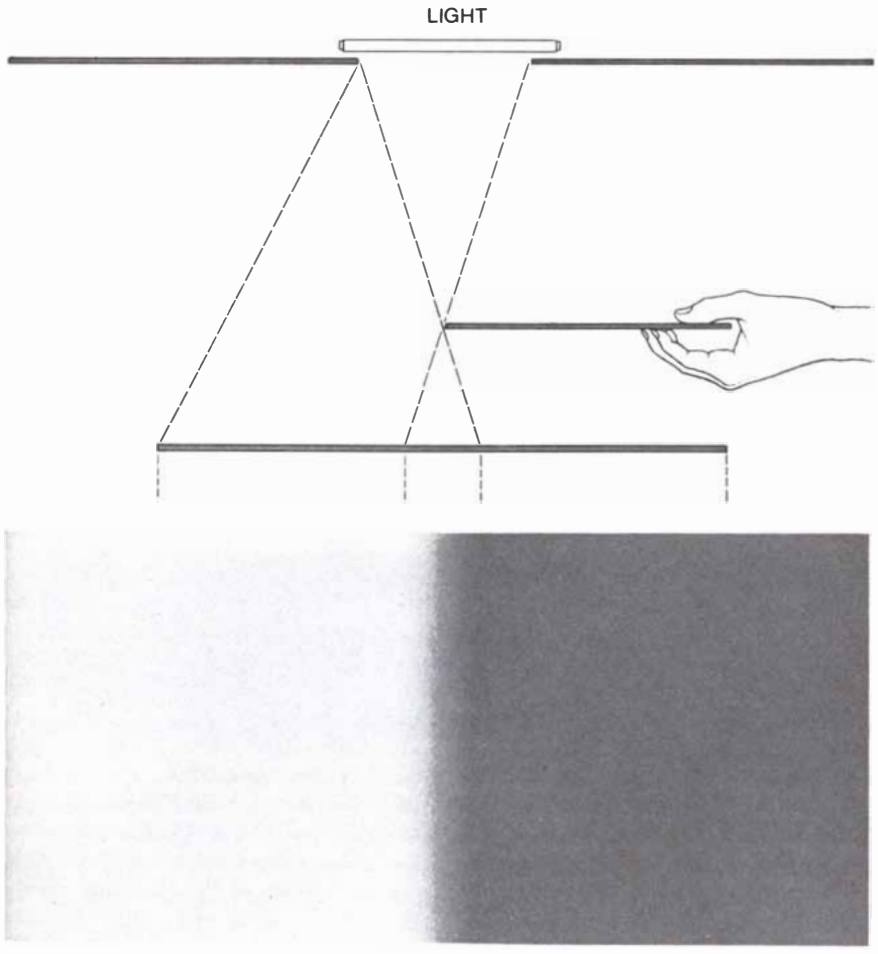
These dark and bright strips, now known as Mach bands, were first reported in the scientific literature some 100 years ago by the Austrian physicist, philosopher and psychologist Ernst Mach. They depend strictly on the distribution of the illumination. Mach formulated a simple principle for the effect: "Whenever the light-intensity curve of an illuminated surface (whose light intensity varies in only one direction) has a concave or convex flection with respect to the abscissa, that place appears brighter or darker, respectively, than its surroundings" [*see bottom illustration on next page*].

The basic effect can be demonstrated by holding an opaque card under an ordinary fluorescent desk lamp, preferably in a dark room. If the shadow is cast on a piece of paper, part of the paper is illuminated by light from the full length of the lamp. Next to the illuminated area is a half-shadow that gets progressively darker until a full shadow is reached. Ideally the distribution of light should be uniformly high in the bright area, uniformly low in the dark area and smoothly graded between the bright and the dark areas [*see top illustration on next page*]. If you now look closely at the edges of the graded half-shadow, you see a narrow bright band at the bright edge and a narrow dark band at the dark edge. These are the Mach bands. Their appearance is so striking that many people will not believe at first that they are only a subjective phenomenon. Some will mistakenly try to explain the appearance of the bands by saying they are the result of multiple shadows or diffraction.

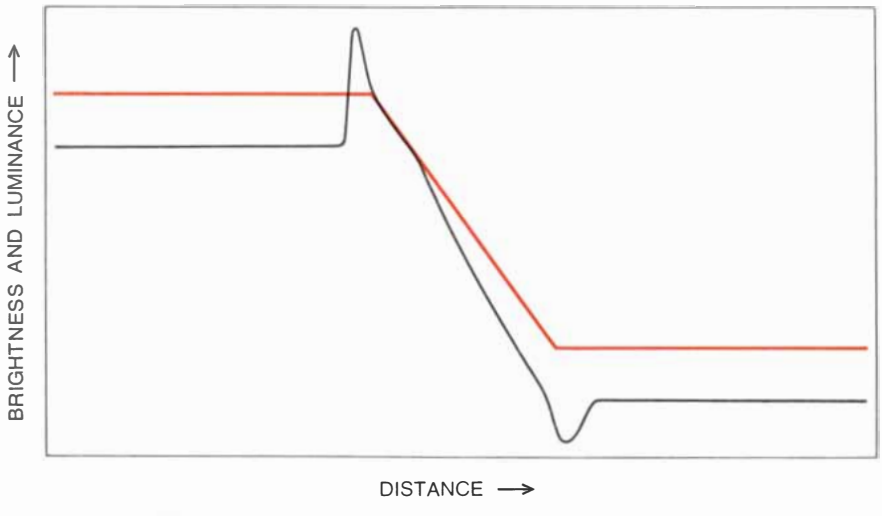
Exact psychophysical measurements of the subjective appearance of Mach bands have been made by Adriana Fiorentini and her colleagues at the National Institute of Optics in Italy. Their technique consists in having an observer adjust an independently variable spot of light to match the brightness of areas in and around the Mach bands. In general they find that the bright band is distinctly narrower and more pronounced than the dark band. The magnitude of the effect, however, varies considerably from person to person.

Since Mach bands delineate contours we expect to see, only a careful observer, or someone who has reason to objectively measure the light distribution at a shadow's edge, is likely to realize that the bands are a caricature of the actual pattern of illumination. Artists of the 19th-century Neo-Impressionist school were unusually meticulous in their observations, and this was reflected in much of their work. A good example is Paul Signac's "Le petit déjeuner." In this painting there are numerous contrast effects in and around the shadows and half-shadows. Particularly striking is how some of the shadows are darkest near their edges and quite light near

NEO-IMPRESSIONIST PAINTER Paul Signac was a meticulous observer of the contrast effects in shadows and half-shadows. On the opposite page is a portion of his "Le petit déjeuner" (1886–1887). Note how the shadow is darker near the unshaded tablecloth and lighter next to the dark matchbox. Similar effects can be found in other shadows. The effects change when the painting is viewed from various distances. The painting is in the Rijksmuseum Kröller-Müller at Otterlo in the Netherlands and is reproduced with its permission.



MACH BANDS can be produced with light from an ordinary fluorescent desk lamp (*upper illustration*). Place a sheet of white or gray paper on the desk and the light about a foot or so above it. Covering the ends of the lamp, which usually are not uniformly bright, may enhance the effect. Turn out the other lights in the room and hold an opaque card an inch or less above the paper. Various positions should be tried for optimum results. Note the narrow bright line and the broader dark line at the outer and inner edges of the half-shadow; these are the Mach bands. The lower illustration is a photograph of a half-shadow produced by the method described. The reproduction of the photograph does not retain all the characteristics of the original because of losses inherent in the reproduction process.



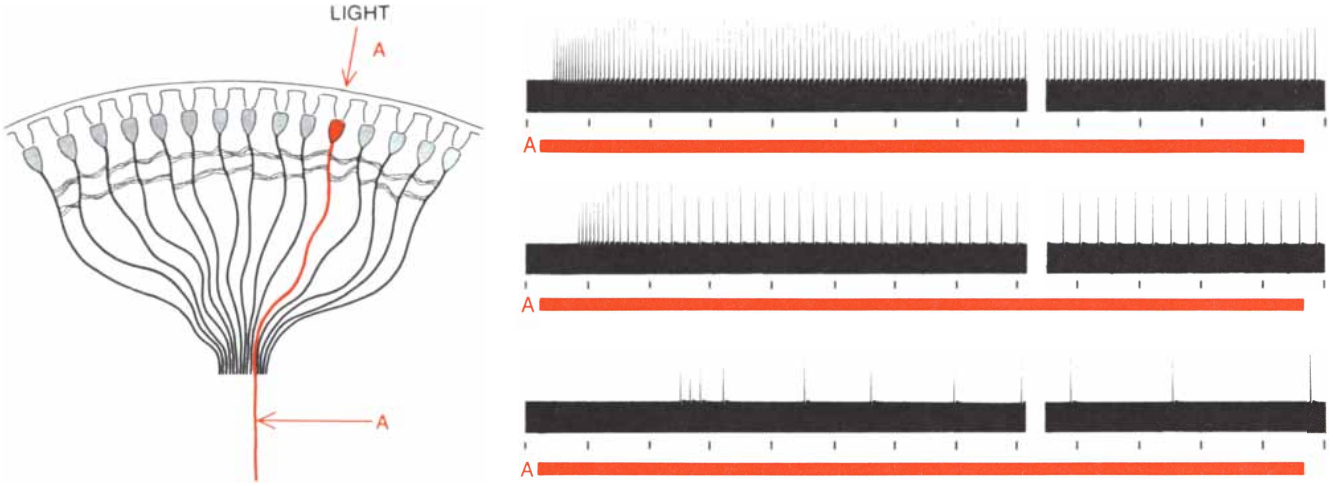
OBSERVED BRIGHTNESS CURVE obtained by psychophysical measurements (*black line*) has two sharp flections, one corresponding to the bright band and the other to the dark band. Measurement of actual luminance (*colored line*) across a half-shadow region reveals that the effect lies in the eye of the beholder and is not an objective phenomenon.

the object casting the shadow. Where Signac saw contrast he painted contrast, whether it was objectively present in the original scene or not. The effects we see in his painting depend of course partly on what Signac painted and partly on how our own eyes respond to contrast. When we view Signac's painting, our own eyes and brain further exaggerate the contrast he painted. As a result the painting appears to have even more contrast than the original scene could have had.

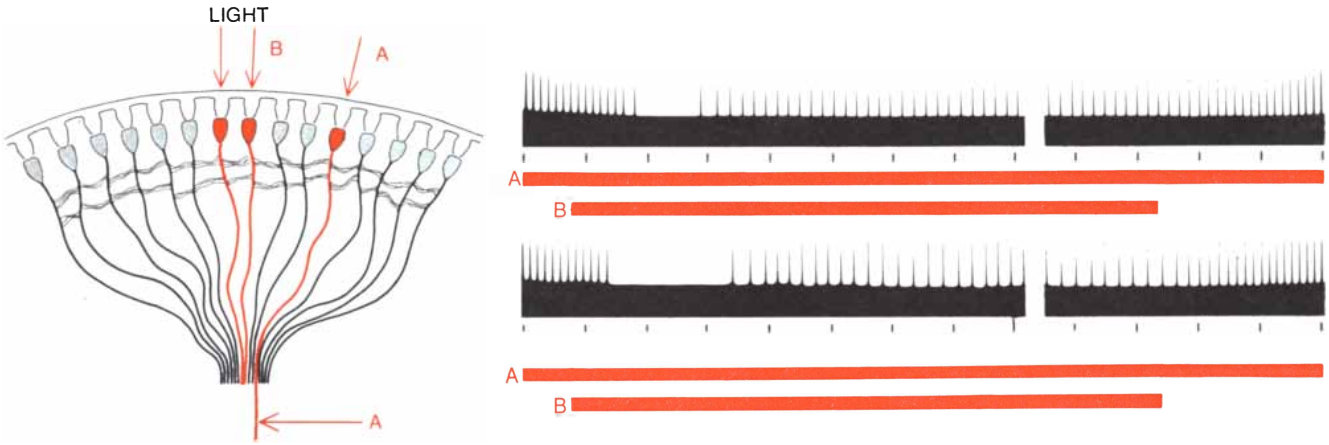
Without precise physical and psychophysical measurements it is difficult to tell how much of the contrast we perceive is objective and how much is subjective. Adding to the confusion is the fact that the subjective Mach bands can seemingly be photographed. All the photograph does, however, is to reproduce with considerable fidelity the original distribution of light in a scene, and it is this distribution of light and dark that gives rise to the subjective Mach bands. Moreover, the photographic process can itself introduce a spurious enhancement of contrast. Edge effects that closely resemble Mach bands can arise as the film is developed. Unlike Mach bands, they are an objective phenomenon consisting of actual variations in the density of the film, and the variations can be objectively measured.

On many occasions scientific investigators have mistaken Mach bands for objective phenomena. For example, shortly after W. K. Röntgen discovered X rays several workers attempted to measure the wavelength of the rays by passing them through ordinary diffraction slits and gratings and recording the resulting pattern on film. Several apparently succeeded in producing diffraction patterns of dark and light bands from which they could determine the wavelength of the X rays. All, however, was in error. As two Dutch physicists, H. Haga and C. H. Wind, showed later, the supposed diffraction patterns were subjective Mach bands.

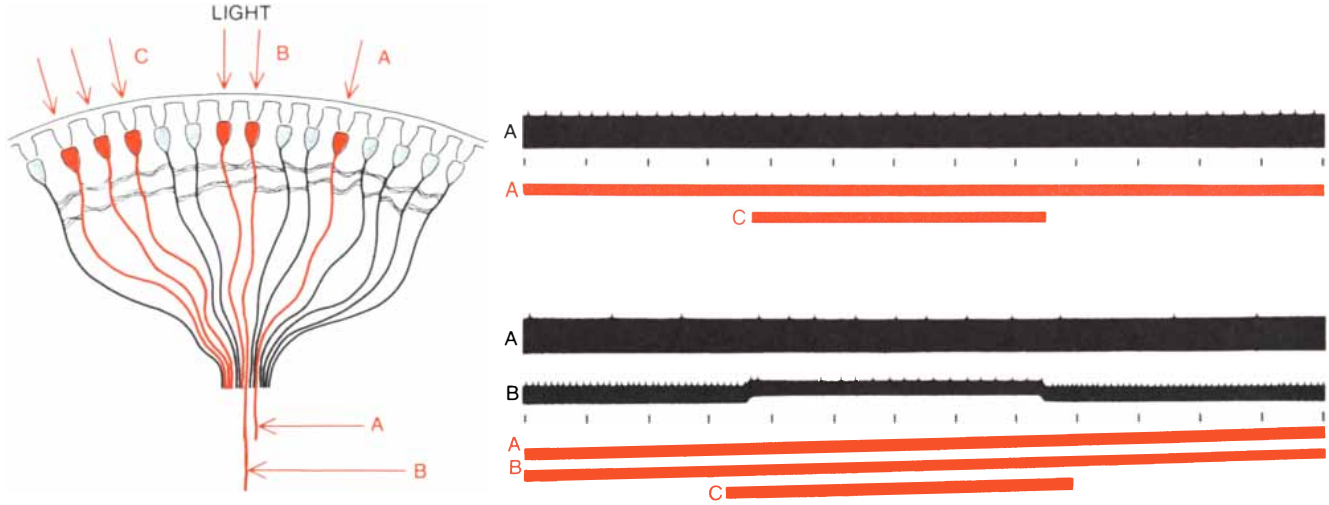
As early as 1865 Mach proposed an explanation of the subjective band effect and other contrast phenomena in terms of opposed excitatory and inhibitory influences in neural networks in the retina and the brain. The means for direct investigation of such neural mechanisms did not become available, however, until the 1920's, when E. D. Adrian, Y. Zotterman and Detlev W. Bronk, working at the University of Cambridge, developed methods for recording the electrical activity of single nerve cells. The basic excitatory-inhibitory principle



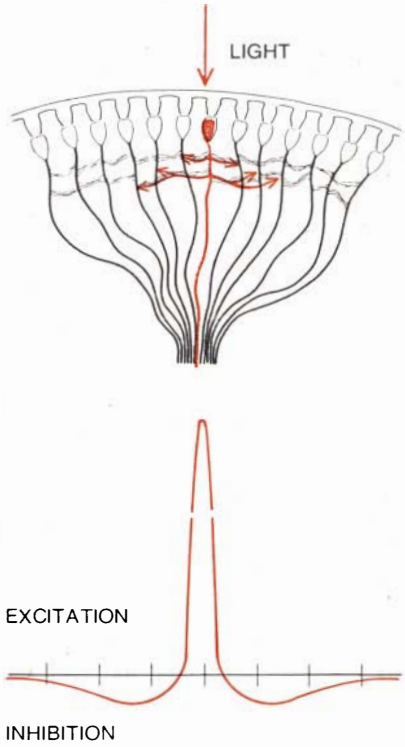
RATE OF DISCHARGE of nerve impulses produced by steady illumination of a single receptor, *A*, in the eye of the horseshoe crab *Limulus* is directly related to the intensity of the light. The nerve fibers from the receptor are separated by microdissection and connected to an electrode from an amplifier and a recorder. The top record shows the response of *A* to steady, high-intensity light. The middle record shows the response to light of moderate intensity, and the lower record the response to low-intensity illumination. Duration of the light signal is indicated by the colored bar. Each mark above the colored bar indicates one-fifth of a second.

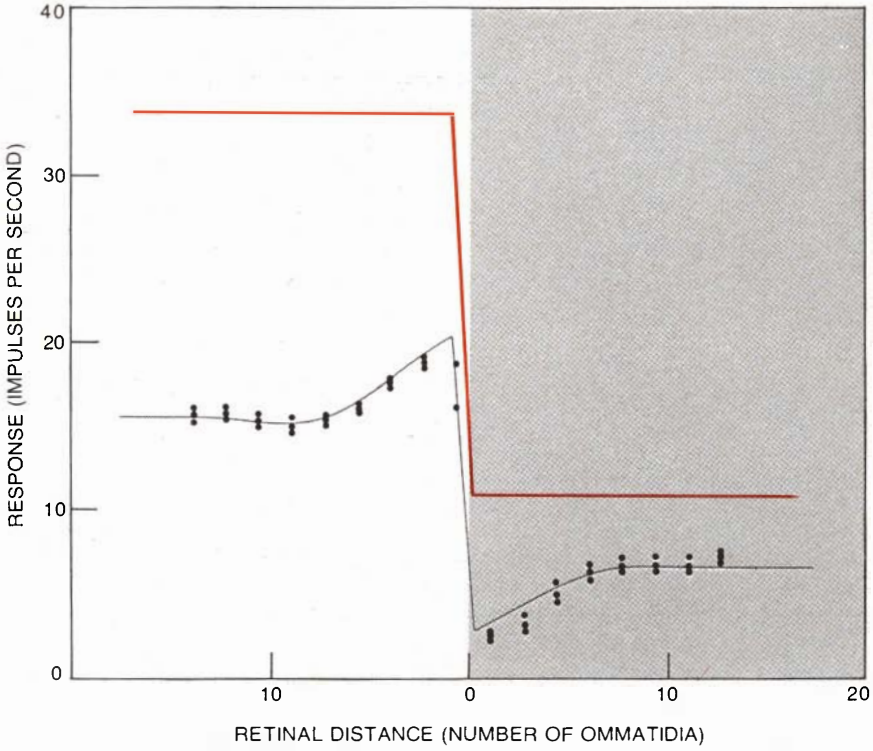


INHIBITION of receptor, *A*, steadily exposed to moderate illumination is produced when neighboring receptors, *B*, are also illuminated. The beginning and the end of the records show the initial and final rate of impulses by *A*. The colored bars indicate duration of light signals. The upper record shows the effects on *A* of moderate-intensity illumination of *B*. The lower record shows the effect on *A* of high-intensity illumination of *B*. The stronger the illumination on neighboring receptors, the stronger the inhibitory effect.

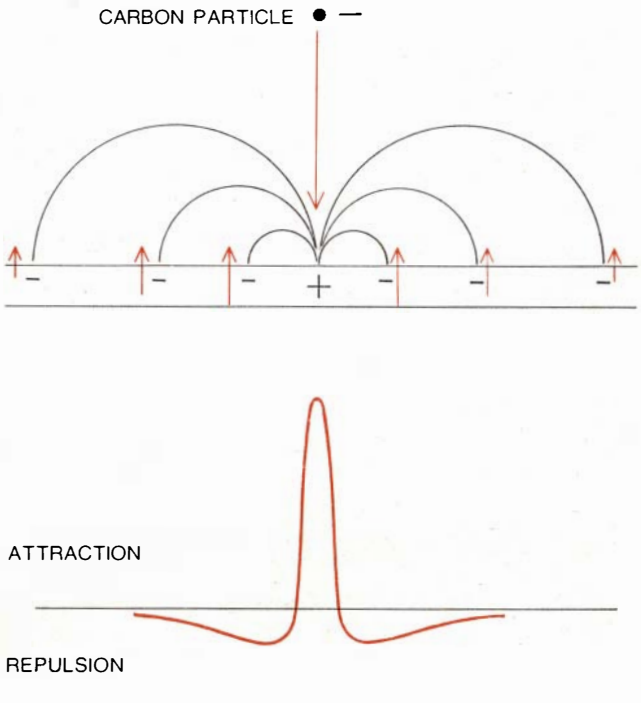


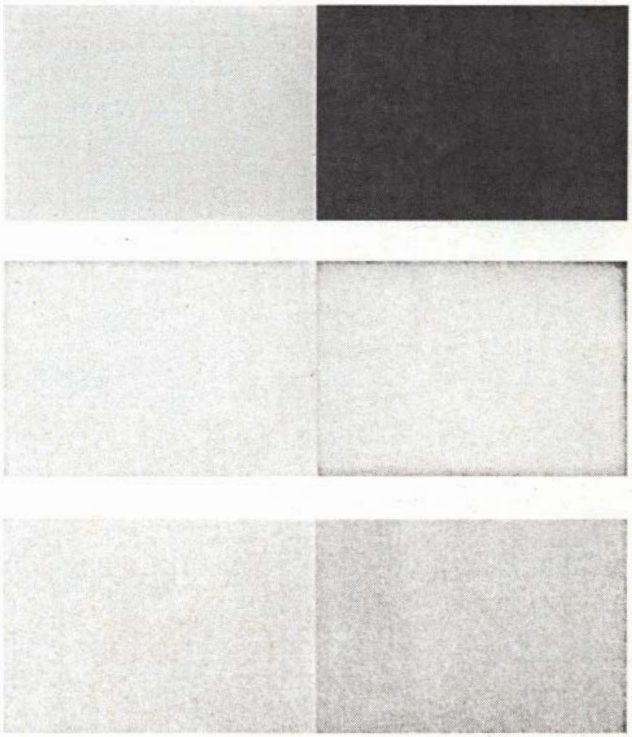
DISINHIBITION of receptor *A* occurs when the inhibition exerted on it by the *B* receptors is partially released by illuminating the large area *C*. The upper record shows that *A*'s activity is not affected when *C* also is illuminated because of the distance between them. The first part of the lower record shows the inhibitory effect of *B* on *A*, then the inhibition of *B* when *C* is illuminated and the concomitant disinhibition of *A*. When the illumination of *C* stops, *B* returns to a higher rate of activity and resumes its inhibition of *A*.





LATERAL INHIBITION in the eye of the horseshoe crab is strongest between receptors a short distance apart and grows weaker as the distance between receptors increases. Below the eye section is a graph of the type of excitatory and inhibitory fields that would be produced by the illumination of a single receptor. The colored line in the graph on the right shows what the retinal response would be to a sharp light-to-dark contour if lateral inhibition did not occur. The points on the graph show responses actually elicited by three scans of the pattern across the receptor in an experiment by Robert B. Barlow, Jr., of Syracuse University. The thin line shows the theoretical responses for lateral inhibition as computed by Donald A. Quarles, Jr., of the IBM Watson Research Center.





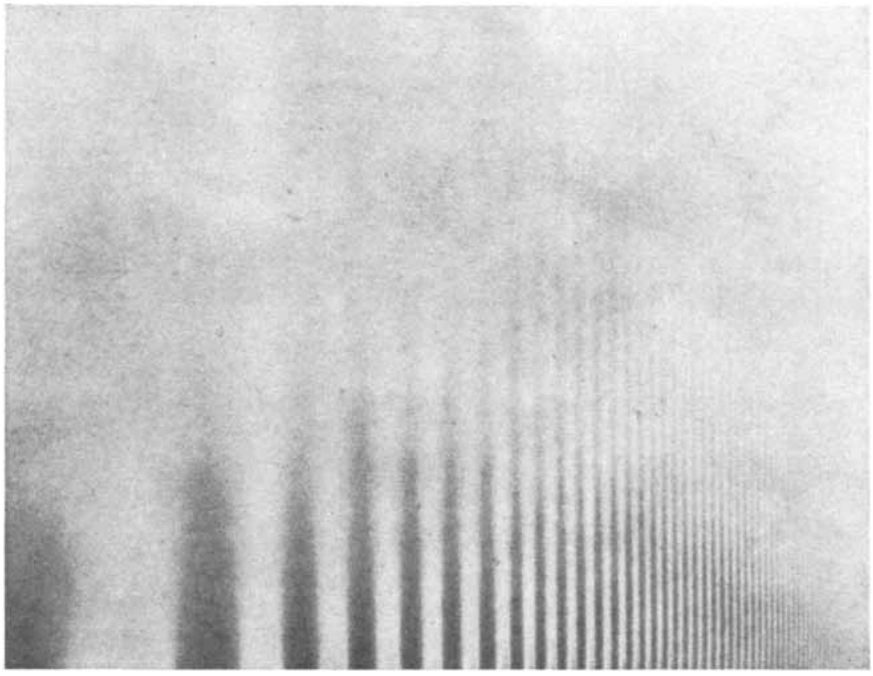
EDGE EFFECT in xerographic copying is the result of the shape of the electrostatic field (which is quite similar to that of the "neural" field in the top illustration) around a single charged point on the xerographic plate (*upper left*). The first panel on the right shows the original pattern. The middle panel shows a Xerox copy of the original. Note how contrast at the edges is greatly enhanced. The bottom panel shows a Xerox copy made with a halftone screen placed over the original so that the pattern is broken up into many dots.

has been demonstrated to be essentially correct in experiments that H. K. Hartline and I, together with our colleagues, have carried out over the past 20 years.

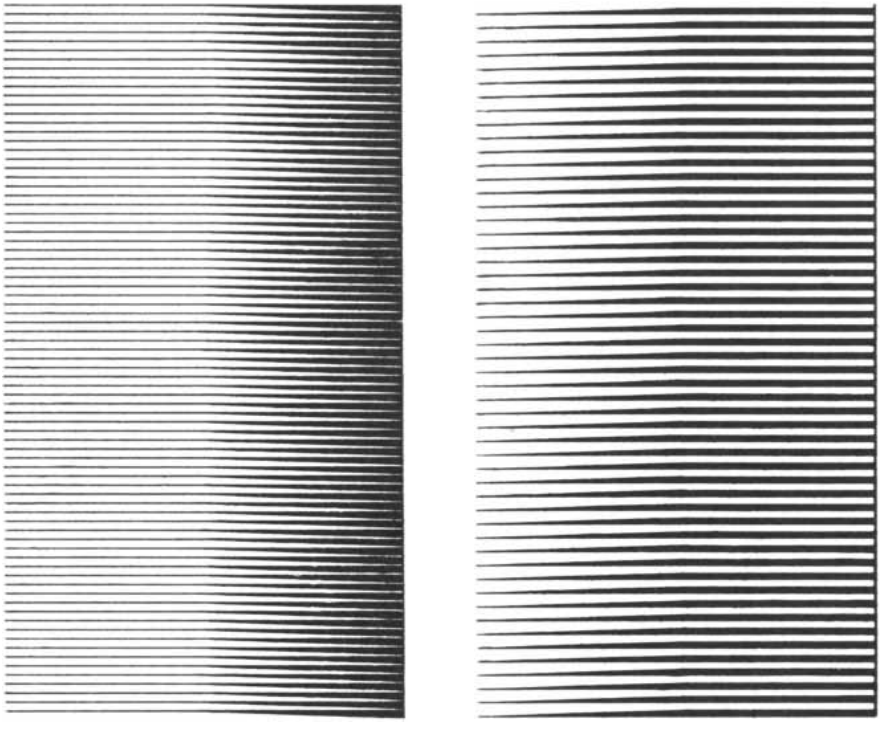
We measured the responses of single neurons in the compound lateral eye of the horseshoe crab *Limulus*. (The animal also has two simple eyes in the front of its carapace near the midline.) The lateral eye of the horseshoe crab is comparatively large (about a centimeter in length) but otherwise it is much like the eye of a fly or a bee. It consists of about 1,000 ommatidia (literally "little eyes"), each of which appears to function as a single photoreceptor unit. Excitation does not spread from one receptor to another; it is confined to whatever receptor unit is illuminated. Nerve fibers arise from the receptors in small bundles that come together to form the optic nerve. Just behind the photoreceptors the small nerve bundles are interconnected by a network of nerve fibers. This network, or plexus, is a true retina even though its function is almost purely inhibitory.

Both the local excitatory and the extended inhibitory influences can be observed directly. A small bundle of fibers from a single receptor is separated by microdissection from the main trunk of the optic nerve and placed on an electrode. In this way the nerve impulses generated by light striking the receptor can be recorded. Weak stimulation produces a low rate of discharge; strong stimulation produces a high rate. These responses are typical of many simple sense organs.

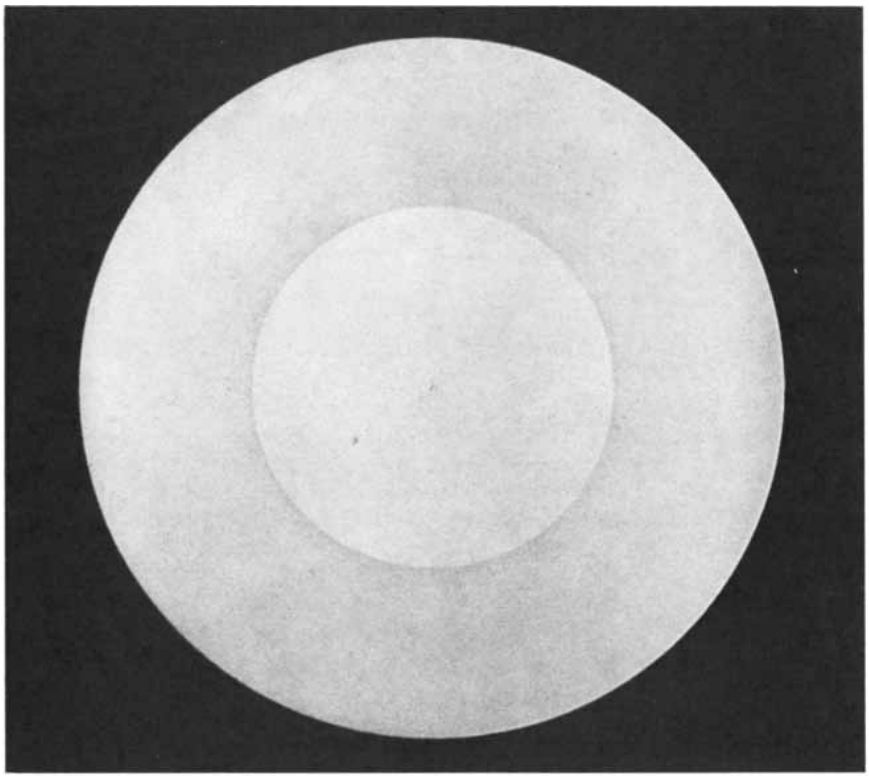
In addition to the excitatory discharge there is a concomitant inhibitory effect. When a receptor unit fires, it inhibits its neighbors. This is a mutual effect: each unit inhibits others and in turn is inhibited by them. The strength of the inhibition depends on the level of activity of the interacting units and the distance between them. In general near neighbors affect one another more than distant neighbors, and the stronger the illumination, the stronger the inhibitory effect. We discovered that such an organization can produce a second-order effect that we call disinhibition. If two sets of receptors are close enough together to interact, they inhibit each other when both sets are illuminated. Now suppose a third set of receptors, far enough away so that it can interact with only one of the two sets of receptors, is illuminated. The activity of the third set will inhibit one set of the original pair, which in turn reduces the inhibition on



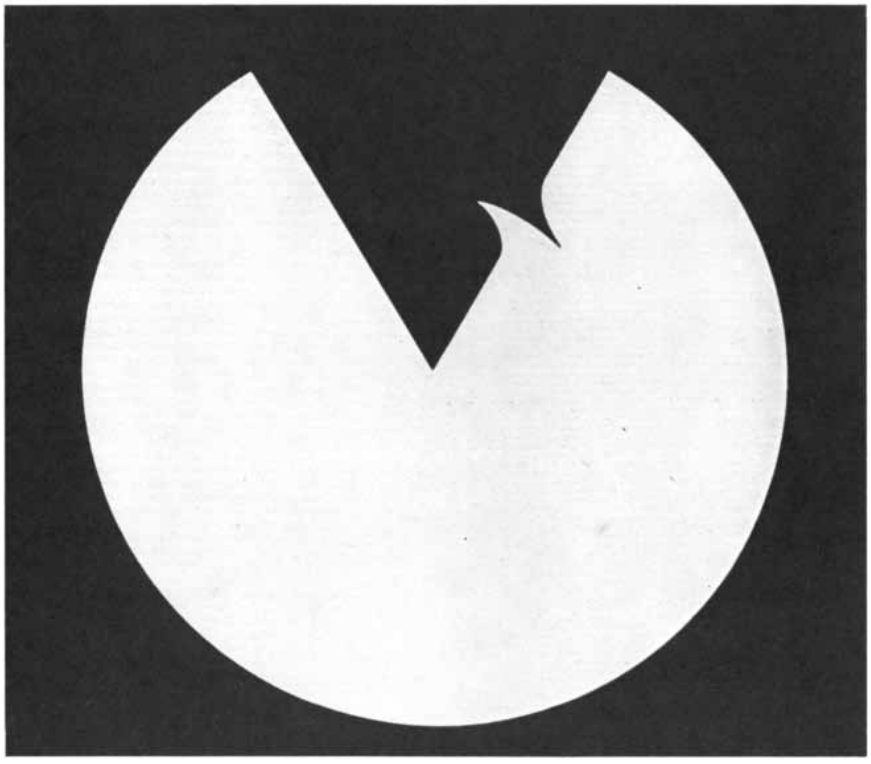
FILTER produced by lateral inhibition at low spatial frequencies and the lack of resolving power of the retina at high spatial frequencies causes intermediate spatial frequencies to be the most distinctly seen. The width of the vertical dark and light bands decreases in a logarithmic sinusoidal manner from the left to the right; the contrast varies logarithmically from less than 1 percent at the top to about 30 percent at the bottom. The objective contrast at any one height in the figure is the same for all spatial frequencies, yet the spatial frequencies in the middle appear more distinct than those at high or low frequencies; that is, the dark lines appear taller at the center of the figure. The effects of changes in viewing distance, luminance, adaptation and sharpness of eye focus can be demonstrated by the viewer.



MACH BAND PHENOMENON created with horizontal lines is shown here. In the illustration at left the black lines are a constant thickness from the left side to the midpoint and then thicken gradually. When the illustration is viewed from a distance, a vertical white "Mach band" appears down the middle. In the illustration at right the horizontal black lines are a constant thickness from the right side to the midpoint and then thin out. When viewed from a distance, the illustration appears to have a vertical black band down the middle.



CRAIK-O'BRIEN EFFECT (this example is known as the Cornsweet illusion) is the result of a specific variation of luminance at the contour, which makes the outer zone appear slightly darker even though it has the same luminance as the inner zone. The effect here is less than in the original because of difficulty in reproducing the actual intensity relations.



RAPID ROTATION of this disk will create the Cornsweet illusion. The white spur creates a local variation near the contour between the two zones that causes the apparent brightness of the inner zone to increase. In the same way the dark spur creates a local variation that causes the outer zone to appear darker. Except in the spur region the objective luminance of the disk when it is rotating is the same in both the inner and the outer region.

the remaining set, thus increasing their rate of discharge [*see bottom illustration on page 93*]. Following the discovery of disinhibition in the eye of the horseshoe crab, Victor J. Wilson and Paul R. Burgess of Rockefeller University found that some increases in neural activity (called recurrent facilitation) that had been observed in spinal motoneurones in the cat were actually disinhibition. Subsequently M. Ito and his colleagues at the University of Tokyo observed a similar type of disinhibition in the action of the cerebellum on Deiter's nucleus in the cat.

The spatial distribution and relative magnitudes of the excitatory and inhibitory influences for any particular receptor unit in the eye of *Limulus* can be represented graphically as a narrow central field of excitatory influence surrounded by a more extensive but weaker field of inhibitory influence [*see top illustration on page 94*].

As Georg von Békésy has shown, the approximate response of an inhibitory network can be calculated graphically by superimposing the graphs for each of the interacting units, each graph scaled according to the intensity of the stimulus where it is centered. The summed effects of overlapping fields of excitation (positive values) and inhibition (negative values) at any particular point would determine the response at that point. In the limit of infinitesimally small separations of overlapping units, this would be mathematically equivalent to using the superposition theorem or the convolution integral to calculate the response. In fact, these inhibitory interactions may be expressed in a wide variety of essentially equivalent mathematical forms. The form Hartline and I used at first is a set of simultaneous equations—one equation for each of the interacting receptor units. Our colleagues Frederick A. Dodge, Jr., Bruce W. Knight, Jr., and Jun-ichi Toyoda have since that time expressed the properties of the inhibitory network in a less cumbersome and more general form: a transfer function relating the Fourier transform of the distribution of the intensity of the stimulus to the Fourier transform of the distribution of the magnitude of the response. This in effect treats the retinal network as a filter of the sinusoidal components in the stimulus, and can be applied equally well to both spatial and temporal variations. The overall filtering effect of the *Limulus* retina is to attenuate both the lowest and the highest spatial and temporal frequencies of the sinusoidal components.

It has long been known that spatial and temporal filtering effects of much the same kind occur in our own visual system. The main characteristics of the spatial "filter" can be seen by viewing the test pattern devised by Fergus W. Campbell and his colleagues at the University of Cambridge [*see top illustration on page 95*].

Even without considering the filter-like properties of neural networks it is possible to see how the subjective Mach bands can be produced by the interaction of narrow fields of excitation and broad fields of inhibition. Near the boundary between the light and dark fields some of the receptors will be inhibited not only by their dimly lit neighbors but also by some brightly lit receptors. The total inhibition of these boundary receptors will therefore be greater than the inhibition of dimly lit receptors farther from the boundary. Similarly, a brightly lit receptor near the boundary will be in the inhibitory field of some dimly lit receptors and as a result will have less inhibition acting on it than brightly lit receptors farther away from the boundary. Because of these differential effects near the boundary the response of the neural network in the *Limulus* retina will show a substantial maximum and minimum adjacent to the boundary even though the stimulus does not have such variations.

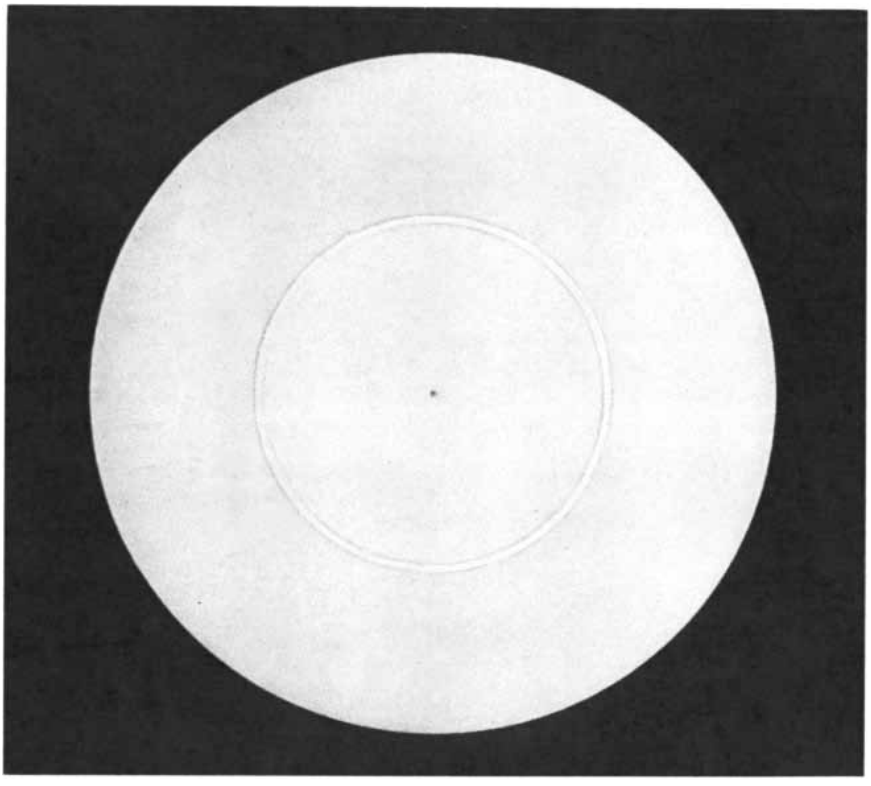
Opposed excitatory and inhibitory influences can mediate some highly specialized functions in higher animals. Depending on how these opposed influences are organized, they can detect motion, the orientation of a line or the difference between colors. No matter how complicated the visual system is, however, the basic contrast effects of the excitatory-inhibitory processes show up. For example, recent experiments by Russell L. De Valois and Paul L. Pease of the University of California at Berkeley show a contour enhancement similar to the bright Mach band in responses of monkey lateral geniculate cells. The simple lateral inhibition that produces contrast effects such as Mach bands may be a basic process in all the more highly evolved visual mechanisms.

Contrast phenomena are by no means found only in the nervous system. Indeed, contrast is found in any system of interacting components where opposed fields of positive (excitatory) and negative (inhibitory) influences exist. Whether the system is neural, electrical, chemical or an abstract mathematical model is irrelevant; all that is needed to produce a contrast effect is a certain distribution of the opposed influences. A familiar example is the contrast effect in xerography. The xerographic process does not reproduce solid black or gray areas very well. Only the edges of extended uniform areas are reproduced unless some special precautions are taken. This failing is inherent in the basic process itself. In the making of a xerographic copy a selenium plate is first electrostatically charged. Where light falls on the plate the electrostatic charge is lost; in dark areas the charge is retained. A black powder spread over the plate clings to the charged areas by electrostatic attraction and is eventually transferred and fused to paper to produce the final copy.

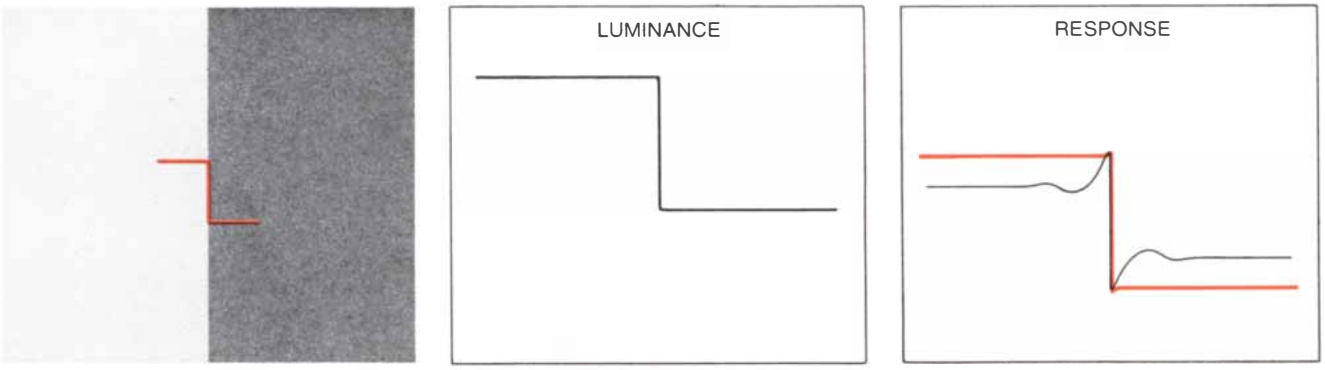
The electrostatic attraction of any point on the plate is determined not by the charge at that point alone but by the integrated effects of the electrostatic fields of all the charges in the neighborhood. Since the shapes of the positive and negative components of the individual fields happen to be very much like the shapes of the excitatory and inhibitory components of neural unit fields in the retina, the consequences are much the same too [*see bottom illustration on page 94*]. Contours are enhanced; uniform areas are lost. To obtain a xerographic copy of the uniform areas one merely has to put a halftone screen over the original. The screen breaks up the uniform areas into many small discontinuities, in effect many contours.

Similar contrast effects are seen in photography and in television. In photography a chemical by-product of the development process at one point can diffuse to neighboring points and inhibit further development there, causing spurious edge effects; in television the secondary emission of electrons from one point in the image on the signal plate in the camera can fall on neighboring points and "inhibit" them, creating negative "halos," or dark areas, around bright spots. The similarity of the contrast effects in such diverse systems is not a trivial coincidence. It is an indication of a universal principle: The enhancement of contours by contrast depends on particular relations among interacting elements in a system and not on the particular mechanisms that achieve those relations.

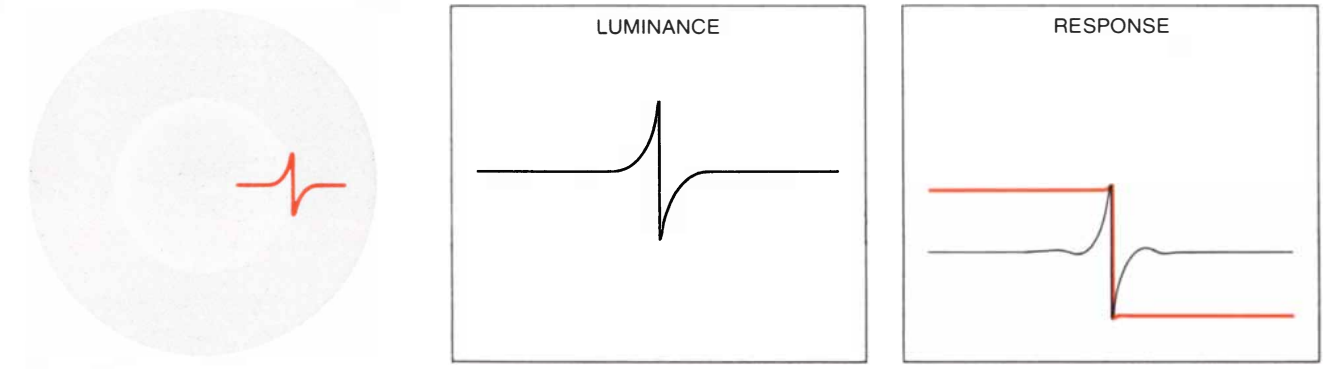
How a contour itself can affect the contrast of the areas it separates cannot be explained quite so easily. This effect of contour on contrast was first investigated by Kenneth Craik of the University of Cambridge and was described in



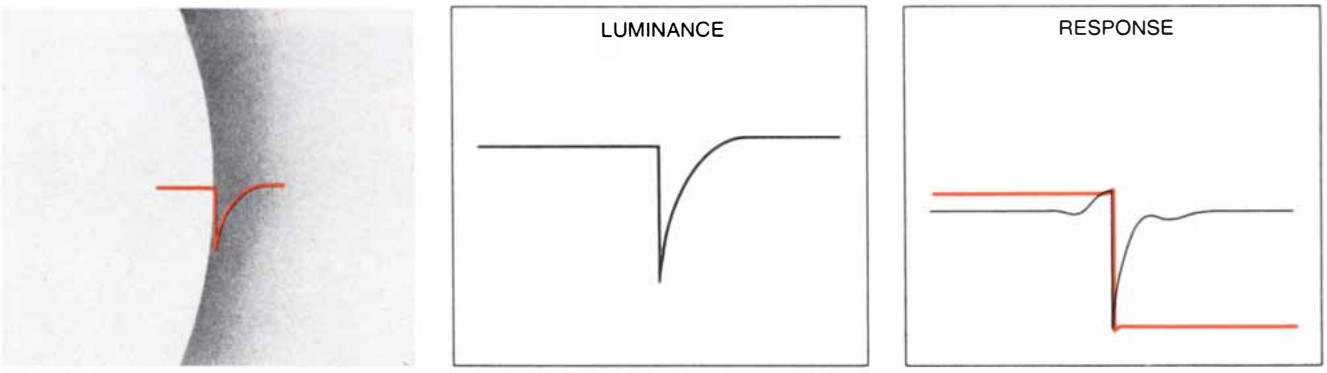
SOURCE OF CRAIK-O'BRIEN EFFECT can be demonstrated by covering the contour with a wire or string. When this is done, the inner and outer regions appear equally bright.



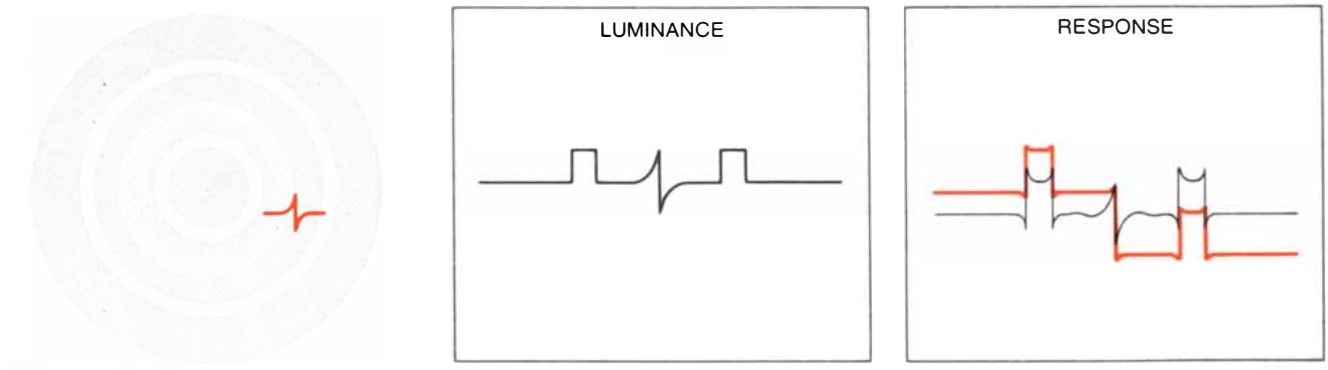
STEP PATTERN of illumination (*left*) also has a step-pattern luminance curve (as measured by a photometer) across the contour. A computer simulation of the response of the *Limulus* eye to the pattern (*black curve at right*) shows a maximum and a minimum that are the result of inhibitory interaction among the receptors. The colored curve at right shows how the pattern looks to a person; the small peak and dip in the curve indicate slight subjective contrast enhancement at the contour known as "border contrast."



LUMINANCE on both sides of the Craik-O'Brien contour is the same but the inside (here simulated) is brighter. The human visual system may extrapolate (*colored curve*) from the maximum and minimum produced by inhibitory processes (*black curve at right*).



DARK SPUR between areas can create brightness reversal. Objectively the area at left of the contour is darker than the area at far right, but to an observer the left side (here simulated) will appear to be brighter than the right side. This brightness reversal agrees with the extrapolation (*colored curve*) from the maximum and minimum produced by inhibitory processes (*black curve at right*).



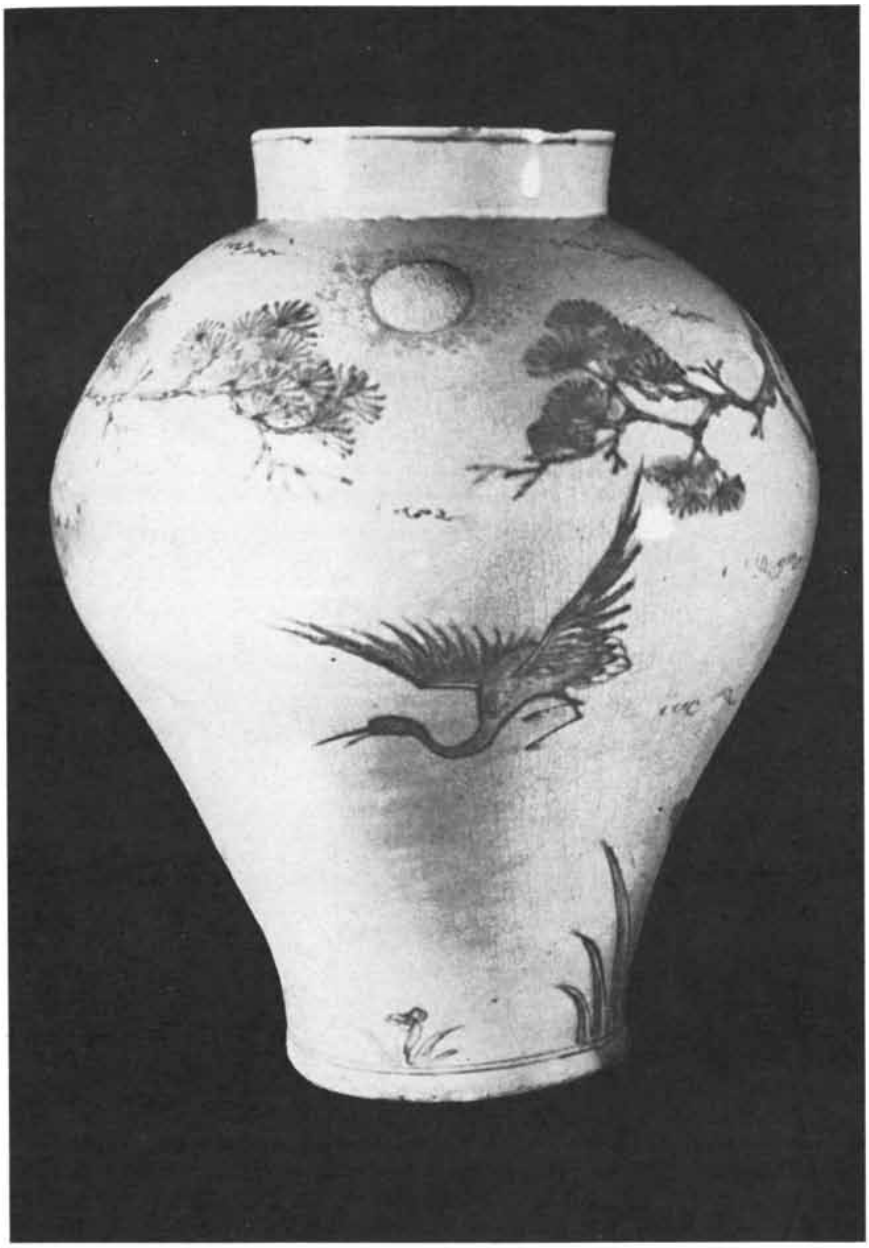
TWO BANDS OF LIGHT of equal intensity are superimposed on backgrounds of equal luminance separated by a Craik-O'Brien contour. The lights add their luminance to the apparent brightness (*colored curve*) and one band appears brighter than the other.

his doctoral dissertation of 1940. Craik's work was not published, however, and the same phenomenon (along with related ones) was rediscovered by Vivian O'Brien of Johns Hopkins University in 1958. The Craik-O'Brien effect, as I shall call it, has been of great interest to neurophysiologists and psychologists in recent years.

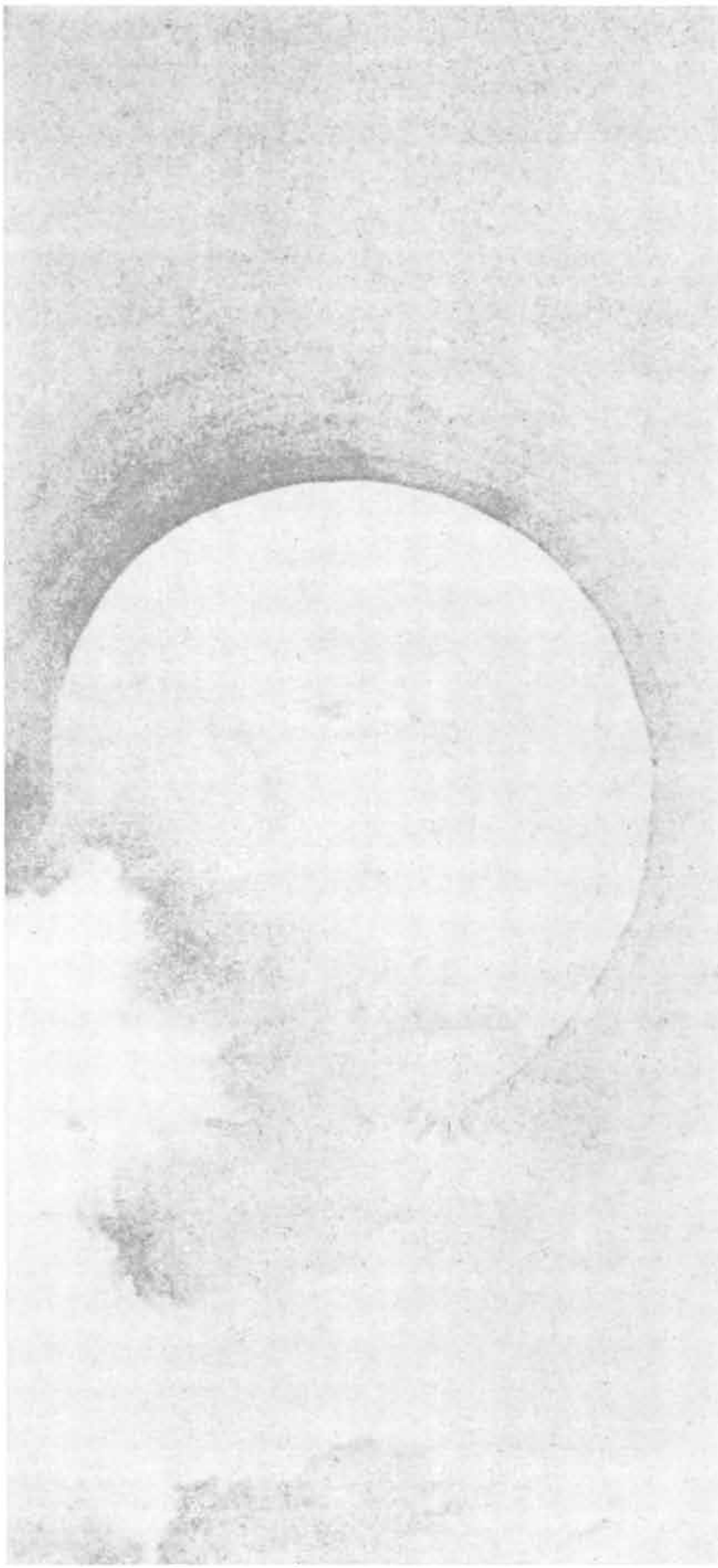
A particular example of this effect, sometimes called the Cornsweet illusion, is produced by separating two identical gray areas with a special contour that has a narrow bright spur and a narrow dark spur [*see top illustration on page 96*]. Although the two uniform areas away from the contour have the same objective luminance, the gray of the area adjacent to the light spur appears to be lighter than the gray of the area adjacent to the dark spur. When the contour is covered with a thick string, the grays of the two areas are seen to be the same. When the masking string is removed, the difference reappears but takes a few moments to develop. These effects can be very pronounced; not only can a contour cause contrast to appear when there actually is no difference in objective luminance but also a suitable contour can cause contrast to appear that is the reverse of the objective luminance.

With the choice of the proper contour a number of objectively different patterns can be made to appear similar in certain important respects [*see illustrations on opposite page*]. It is reasonable to assume that in all these cases the dominant underlying neural events are also similar. With the mathematical equation for the response of a *Limulus* eye one can calculate the neural responses to be expected from each type of pattern when processed by a simple inhibitory network. When this is done, one finds that the calculated responses are all similar to one another. Each has a maximum on the left and a minimum on the right. Furthermore, there is a certain similarity between the calculated neural response and the subjective experience of a human observer viewing the patterns: where the computed response has a maximum, the pattern appears brighter on that side of the contour; where the computed response has a minimum, the pattern appears darker on that side of the contour. Indeed, merely by extending a line from the maximum out to the edge of that side of the pattern and a line from the minimum out to the edge of that side of the pattern one obtains a fair approximation to the apparent brightness. This correspondence suggests that opposed excitatory and inhibitory influences in neural networks of our visual systems are again partly responsible for creating the effect. Even so, much would remain to be explained. Why should the influence of the contour be extended over the entire adjacent area rather than just locally? And why do three distinctly different stimuli, when used as contours, produce much the same subjective result?

The answer to both of these questions may be one and the same. Communication engineers have experimented with a number of sophisticated means of data compression to increase the efficiency of transmitting images containing large amounts of redundant information. For example, if a picture is being transmitted, only information about contours need be sent; the uniform areas between contours can be restored later by computer from information in the amplitudes of the maxima and minima at the contours. By the same token signals from the retina may be "compressed" and the redundant information extrapolated from the maximum and the minimum in the neural response. Such a process, which was postulated by Glenn A. Fry



KOREAN VASE from the 18th century provides an excellent example of the effect of a dark spur between areas. The moon appears to be brighter than the sky directly below it, but the actual luminance is just the reverse. If only a portion of the moon and an equal portion of the sky about one moon diameter below it are viewed through two identical small holes in a paper so that the dark contour is masked, the moon appears darker than the sky.



of Ohio State University many years ago, could explain the Craik-O'Brien effect.

What the actual mechanisms might be in our visual system that could "decode" the signals resulting from data compression by the retina and "restore" redundant information removed in the compression are empirical problems that have not yet been directly investigated by neurophysiologists. The problem as I have stated it may even be a will-o'-the-wisp; it is possible that there is no need to actually restore redundant information. The maximum and minimum in the retinal response may "set" brightness discriminators in the brain, and provided that there are no intervening maxima and minima (that is, visible contours) the apparent brightness of adjacent areas would not deviate from that set by the maximum or the minimum.

Some evidence that apparent brightness is actually set by the maximum and minimum at a contour or discontinuity and is then extrapolated to adjacent areas can be found in experiments conducted by L. E. Arend, J. N. Buehler and Gregory R. Lockhead at Duke University. They worked with patterns similar to those that create the Craik-O'Brien effect. On each side of the contour they produced an additional band of light. They found that the difference in apparent brightness between each band of light and its background depended only on the actual increment in luminance provided by the band, but that the apparent brightness of the two bands in relation to each other was determined by the apparent brightness of the background. For example, if two bands of equal luminance are superimposed on two backgrounds of equal luminance that are separated by a Craik-O'Brien contour, one of the bands of light will appear brighter than the other [*see bottom illustration on page 98*]. A number of related phenomena, in which contrast effects are propagated across several adjacent areas, are under investigation by Edwin H. Land and John H. McCann at the Polaroid Corporation. These experi-

JAPANESE INK PAINTING, "Autumn Moon" by Keinen, has a moon that objectively is only very slightly lighter than the sky. Much of the difference in apparent brightness is created by the moon's contour. The extent of the effect can be seen by covering the moon's edge with string. The painting, made about 1900, is in the collection of the late Akira Shimazu of Nara in Japan.

ments lend further support to the general view outlined here.

Of course, the human visual system is far too complex for the simple notion that apparent brightness is determined by difference at contours to be the whole story. Nonetheless, the general idea contains at least the rudiments of an explanation that is consistent with known physiological mechanisms and with the observed phenomena. Several entirely different distributions of illumination may look much the same to the human eye simply because the eye happens to abstract and send to the brain only those features that the objectively different patterns have in common. This type of data compression may be a basic principle common to many different kinds of neural systems.

Even if the cause of the Craik-O'Brien effect is in doubt, the effect itself is incontrovertible. Although the effects of contour on perceived contrast are relatively new to the scientific community, the same effects have long been known to artists and artisans. One can only speculate on how the effects were discovered. Very likely they emerged in some new artistic technique that was developed for another purpose. Once such a technique had been perfected, it doubtless would have persisted and been handed down from generation to generation. Furthermore, following the initial discovery the technique would probably have been applied in other media. In any event such techniques date back as far as the Sung dynasty of China (A.D. 960–1279), and they are still employed in Oriental art. For example, in a Japanese ink painting made about 1900 a single deft stroke of the brush greatly increases the apparent brightness of the moon [*see illustration on opposite page*]. If the contour is covered with a piece of string, the apparent brightness of the moon diminishes and that area is seen to be very little brighter than its surround.

A similar effect is found in a scene on an 18th-century Korean vase [*see illustration on page 99*]. Here the moon is actually darker than the space below it. Measurements of a photograph of the vase with a light meter under ordinary room lights showed that the luminance of the moon was 15 foot-lamberts and the space one moon diameter below was 20 foot-lamberts. The contour effect is so strong that the apparent brightness of the two areas is just the reverse of the objective luminance.

The contour-contrast effect can be produced on a ceramic surface by still another technique. This technique was developed more than 1,000 years ago in the Ting white porcelain of the Sung dynasty and in the northern celadon ceramics of the same period. In the creation of the effect a design was first incised in the wet clay with a knife. The cut had a sharp inner edge and a sloping outer edge. The clay was then dried and covered with a white glaze. The slightly creamy cast of the glaze inside the cuts produces the necessary gradient to create the Craik-O'Brien effect. The result is that the pattern appears slightly brighter than the surround [*see illustration above*]. Since the effect depends on variations in the depth of the translucent monochrome glaze, it is much more subtle than it is in the Japanese painting and in the Korean vase. But then subtlety and restraint were characteristic of the Sung ceramists.

CHINESE TING YAO SAUCER is an example of the famous Ting white porcelain produced in the Sung dynasty of about A.D. 1000. Although the entire surface is covered with only a single creamy white glaze, the incised lotus design appears brighter than the background because of the incisions, which have a sharp inner edge and a graded outer edge, producing exactly the kind of contour that creates an apparent difference in brightness.

These examples of the effects of contrast and contour from the visual sciences and the visual arts illustrate the need for a better understanding of how elementary processes are organized into complex systems. In recent years the discipline of biology has become increasingly analytical. Much of the study of life has become the study of the behavior of single cells and the molecular events within them. Although the analytic approach has been remarkably productive, it does not come to grips with one of the fundamental problems facing modern biological science: how unitary structures and elementary processes are organized into the complex functional systems that make up living organs and organisms. Fortunately, however, we are not faced with an either-or choice. The analytic and the organic approaches are neither incompatible nor mutually exclusive; they are complementary, and advances in one frequently facilitate advances in the other. All that is required to make biology truly a life science, no matter what the level of analysis, is to occasionally adopt a holistic or organic approach. It is probably the elaborate organization of unitary structures and elementary processes that distinguishes living beings from lifeless things.

Reads credit cards.

Read all about them.

The innovation you see at the left is our new electronic card reader. It scans and helps verify credit in seconds. It may be used by itself, or as part of a complete point-of-sale transaction terminal. In fact, we have technology for all types of card or tag readers. And we have electronic sub systems for digital data processing and transmission including modems, and liquid crystal displays.

In the musical department, we've developed the "brain" of an electronic organ. Mounted on an 11″ x 17″ board, we built a musical mini-computer that converts sound waves into digital signals, stores them in memory, then recalls them as sound waves. Truer than any other electric organ in the world. And at a fraction of what it would cost with conventional electronics.

You'll find us in textiles as well. With an electronically controlled knitting machine that changes fabric patterns faster than the fashion industry can change its mind (or a girl can change her outfit). A computer tape does it. In seconds.

Impressed? Keep reading.

We work in the automotive

Reads music.

Reads fashion news.

field, too. For example, our automatic braking system (known as Skid-Trol™) prevents big rigs from jackknifing, even under the worst road conditions.

Another field is utilities. The dispatch control center that we built for the Philadelphia Electric Company gives an instantaneous picture of the entire bulk power complex. Plus a forecast of load factors during the next 15 minutes.

And, of course, we're suppliers to the business machine manufacturers—we're the world's leading supplier of MOS/LSI devices for electronic calculators.

These—and many more.

When it comes to processing, storing, sensing, displaying or transmitting data, nobody can match our experience. Or capabilities. Write us in Anaheim, California 92803. Or phone 714/632-4195.

Maybe we can improve your reading habits and save you money besides.

North American Rockwell Electronics Group

Psychological Factors in Stress and Disease

A new technique separates the psychological and physical factors in stressful conditions. In studies with rats the psychological factors were the main cause of stomach ulcers and other disorders

by Jay M. Weiss

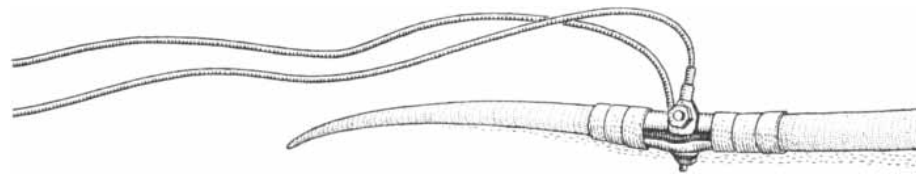
One of the most intriguing ideas in medicine is that psychological processes affect disease. This concept is not new; it dates back to antiquity, and it has always been controversial. To counter those skeptics who believed that "no state of mind ever affected the humors of the blood," Daniel Hack Tuke, a noted 19th-century London physician, compiled an exhaustive volume, *Illustrations of the Influence of the Mind on the Body*. He concluded:

"We have seen that the influence of the mind upon the body is no transient power; that *in health* it may exalt the sensory functions, or suspend them altogether; excite the nervous system so as to cause the various forms of convulsive action of the voluntary muscles, or to depress it so as to render them powerless; may stimulate or paralyze the muscles of organic life, and the processes of Nutrition and Secretion—causing even death; that *in disease* it may restore the functions which it takes away in health, reinnervating the sensory and motor nerves, exciting healthy vascularity and nervous power, and assisting the *vis medicatrix Naturae* to throw off disease action or absorb morbid deposits." Through the years many other individuals have voiced their belief in the importance of psychological factors in disease, and they have carved out the field known as psychosomatic medicine. It is a field filled with questions, and we are still seeking better evidence on the role of the psychological factors.

Our ability to determine the influence of psychological factors on disease entered a new phase recently with the application of experimental techniques. Formerly the evidence that psychological factors influence disease came from the observations of astute clinicians who noted that certain psychological conditions seemed to be associated with particular organic disorders or with the increased severity of disorders. But such evidence, no matter how compelling, is correlational in nature. Although a psychological characteristic or event may coincide with the onset or advance of a disorder, one cannot be certain that it actually has any effect on the disease process; the psychological event may simply occur together with the disease or may even be caused by the disease. In addition it is always possible that the apparent correlation between the psychological variable and the disease is spurious; that among the myriad of other factors in the physical makeup of the patient and in his life situation lies a different critical element the observer has failed to detect. Such considerations can be ruled out only by accounting for every possible element in the disease process, which obviously is impractical.

The development of experimental techniques for inducing disease states in animals has made the task of determining whether or not psychological factors affect disease much easier. When an investigator can establish conditions that will cause a pathology to develop in experimental animals in the laboratory, he does not have to wait for the disease to arise and then attempt to determine if a particular factor had been important; instead he can introduce that factor directly into his conditions and see if it does indeed affect the development of the disease. Moreover, the use of experimental procedures enables the investigator to deal with the numerous other variables that might influence the disease. Even so, the investigator certainly cannot regulate or even be aware of all the variables that affect a disease. Such variables will, however, be distributed randomly throughout the entire population of experimental subjects. When the experimenter applies some treatment to one randomly selected group of subjects and not to another, he knows that any consistent difference between the treatment group and the control group will have been caused by the experimental treatment and not by the other variables, since those variables are distributed randomly throughout both groups.

Recently I have been studying the influence of psychological factors on certain experimentally induced disorders, particularly the development of gastric lesions or stomach ulcers. As the prob-



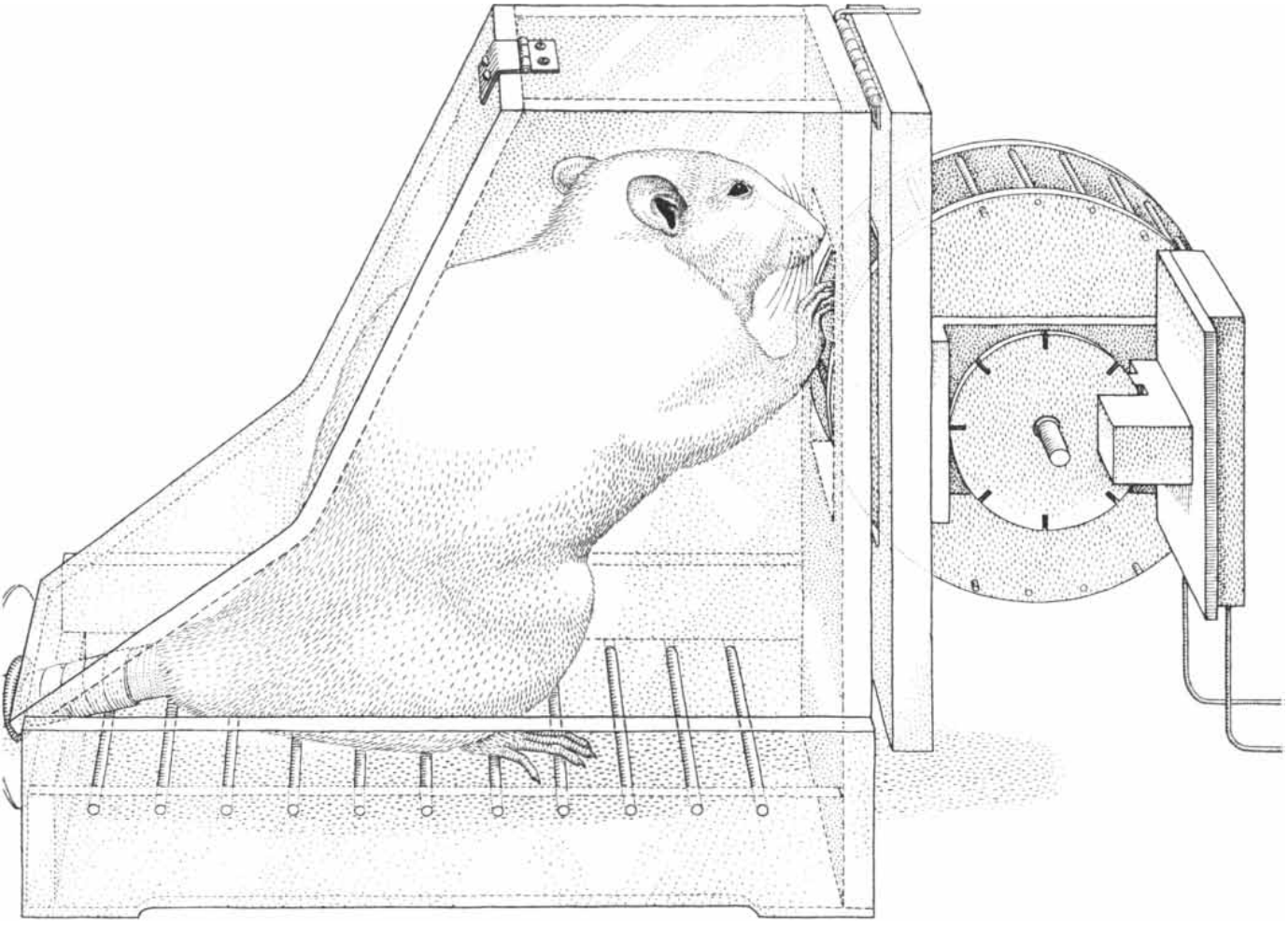
TRANSPARENT-PLASTIC CHAMBER is one type of apparatus used in stress ulcer experiments with rats. The rat is housed in the chamber for the entire 48-hour stress session.

lem of gastrointestinal ulcers has become more widespread (the disease currently causes more than 10,000 deaths each year in the U.S. and afflicts one out of every 20 persons at some point in their life) pathologists have continued to refine techniques of studying ulceration experimentally. Within the past 15 years, beginning with the pioneering work of Serge Bonfils and his associates at the Institut d'Hygiène in Paris, investigators have discovered that experimental animals can develop gastric lesions under stressful environmental conditions. The finding that lesions can be induced by manipulating an animal's external environment opened the way for experimental study of the psychological conditions that are brought about by stressful environmental events.

The experimental techniques had to be refined still further, however, for studying the influence of psychological variables. The reason is that when experimental animals are exposed to an environmental stressor (a stressor is a stress-inducing agent), the effects of psychological variables may be confounded with the effects of the physical stressors. For example, suppose rats are made to swim for an hour in a tank from which they cannot escape, and these animals then develop organic pathology whereas control animals (those not exposed to the swim stressor) show no pathology at all. Although it is evident that the pathology was induced by the swim stressor, how can we assess the role of any psychological factor in producing this pathology? Certainly the pathology might have been affected by the fear the animal experienced, by its inability to escape, by the constant threat of drowning. But what about the extraordinary muscular exertion the stressful situation required, with its attendant debilitation and exhaustion of tissue resources? Clearly in such an experiment it is not possible to determine if psychological variables influenced the development of pathology, since the pathology might have been due simply to the direct impact of the swim stressor itself. Thus in order to study the role of psychological factors one must devise a means of assessing the importance of psychological variables apart from the impact of the physical stressor on the organism.

The method I have used is to expose two or more animals simultaneously to exactly the same physical stressor, with each animal in a different psychological condition. I then look to see if a consistent difference results from these conditions; such a difference must be due to psychological variables since all the animals received the same physical stressor. To illustrate this technique, let us consider an experiment on the effects of the predictability of an electric shock on ulceration.

Two rats received electric shocks simultaneously through electrodes placed on their tail, while a third rat served as a control and received no shocks. Of the two rats receiving shocks, one heard a beeping tone that began 10 seconds before each shock. The other rat also heard the tone, but the tone sounded randomly with respect to the

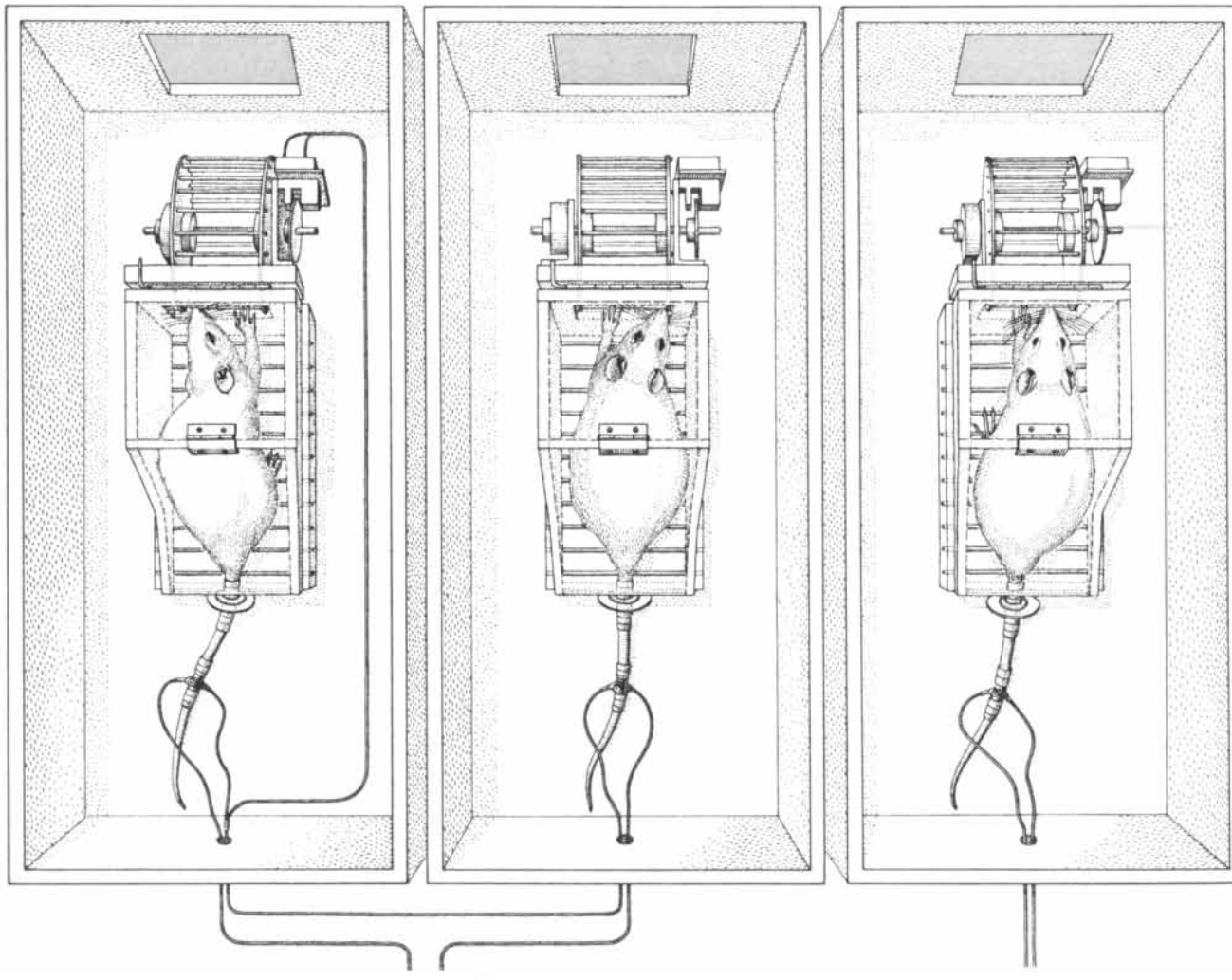


Shocks are delivered through electrodes attached to the tail. A disk secured to the tail by a piece of tubing prevents the rat from pulling its tail into the chamber and biting off the electrodes. The rat receives no food during the experiment, but water is available.

shock. Thus both these animals received the same shocks, but one could predict when the shocks would occur whereas the other could not. Since the physical stressor was the same for the two animals, any consistent difference between them in the amount of ulceration would be the result of the difference in the predictability of the stressor, the psychological variable being studied.

This raises a most important point: If such experiments are to be valid, one must be sure that the physical stressor, in this case the electric shock, is the same for all animals in the test. When these studies were begun, the standard way of administering electric shock to experimental rats was to place them on a grid floor whose bars were electrically charged. That method of delivering an electric shock was clearly inadequate for the present experiments, since rats can lessen the shock on a grid floor by changing their posture or can even terminate the shock completely by jumping. The experimenter is faced with a serious problem if one group of rats is able to perform such maneuvers more effectively than another group. In the predictability experiment, for example, the rats that are able to predict shock would surely have been able to prepare for such postural changes more effectively than the rats that were unable to predict shock, so that the groups would have differed with respect not only to the predictability of the stressor but also to the amount of shock received. That would essentially invalidate the experiment. The tail electrode, which is used in all the experiments discussed here, was developed specifically to avoid the possibility of unequal shock. Because the electrode is fixed to the tail, the rat cannot reduce or avoid the shock by moving about. In addition, with the fixed electrode it is possible to wire the electrodes of matched animals in series, so that both animals are part of the same circuit. Thus all the shocks received by the matched subjects are equal in duration and have an identical current intensity, which appears to be the critical element in determining the discomfort of shock.

As was to be expected, the control rats that received no shock developed very little gastric ulceration or none. A



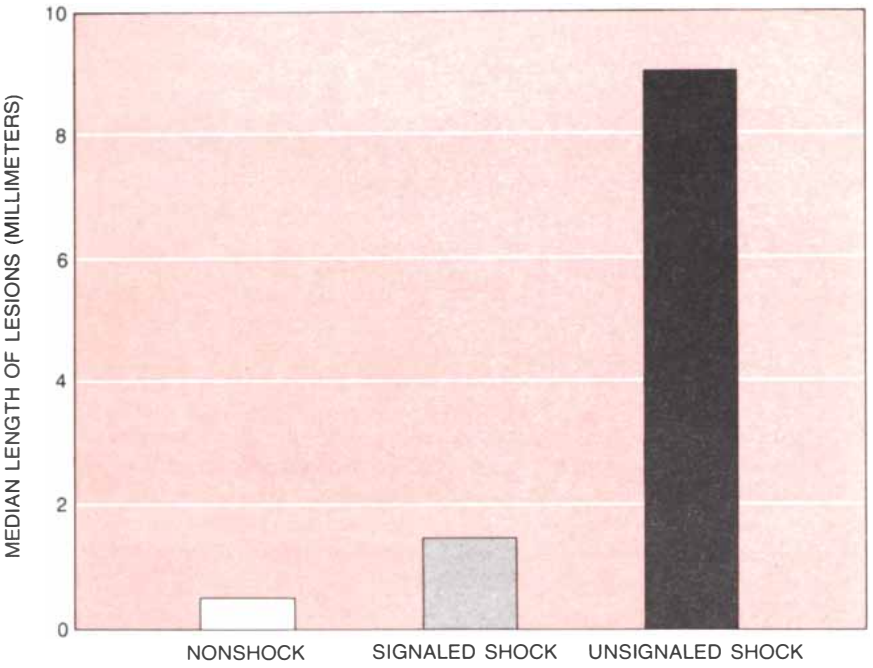
BASIC TRIPLET PARADIGM for the ulceration experiments consists in placing three rats, matched for weight and age, in individual soundproof compartments with one-way-mirror windows. Each rat is prepared in exactly the same way and then randomly assigned to one of the three experimental conditions: avoidance-escape, yoked and control. In this illustration the rat on the left is the avoidance-escape subject. It can terminate the programmed shock by turning the wheel. Moreover, turning the wheel between shocks will postpone the shock. The rat in the center is electrically wired in series to the first rat, so that when the first rat receives a shock, the yoked rat simultaneously receives a shock of the same intensity and duration. The actions of the yoked rat do not affect the shock sequence. The electrodes on the tail of the control rat on the right are not connected, and this rat does not receive shocks at any time. At the end of the experimental session the rats are sacrificed and the length of their gastric lesions is measured.

striking result of the experiment was that rats able to predict when the shocks would occur also showed relatively little ulceration, whereas those that received the same shocks unpredictably showed a considerable amount of ulceration [*see top illustration at right*]. In short, the results demonstrated clearly that the psychological variable of predictability, rather than the shock itself, was the main determinant of ulcer severity.

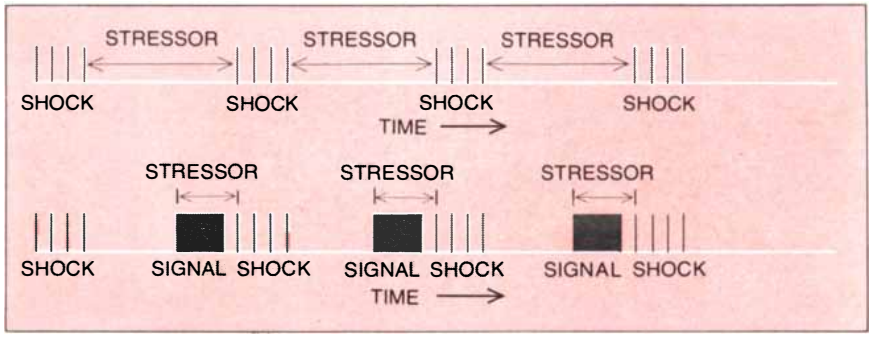
Even though some rats in the foregoing experiment could predict the shock, they were helpless in that they could not avoid the shock. How will stress reactions, such as gastric ulcers, be altered if an animal has control over a stressor instead of being helpless? A recent series of experiments conducted in our laboratory at Rockefeller University has yielded a considerable amount of new information on effects of coping behavior, and also has given us some insight into why these effects arise.

To study the effects of coping behavior, three rats underwent experimental treatment simultaneously, just as in the predictability experiment. Two of the rats again received exactly the same shocks through fixed tail electrodes wired in series, while the third rat served as a nonshock control. In the coping experiments, however, the difference between the two rats receiving shock was not based on predictability (all matched shocked rats in these experiments received the same signals) but rather was based on the fact that one rat could avoid or escape the shock whereas the other could not do anything to escape it.

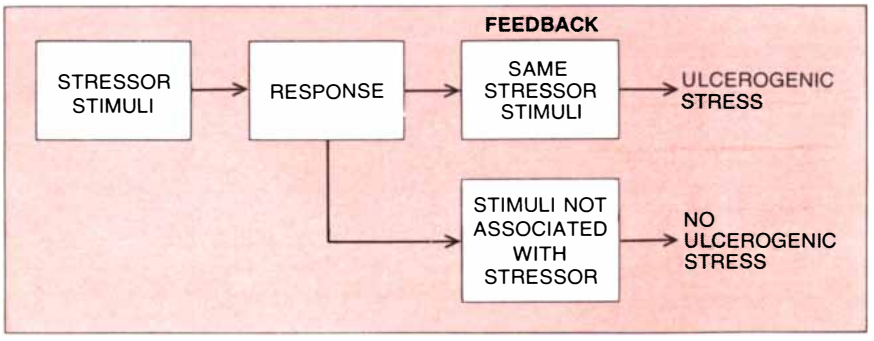
Since albino rats tend to lose weight when stressed, at first I simply measured the effects of avoidance-escape versus helplessness in terms of changes in the rate of weight gain. In this experimental arrangement one rat of each shocked pair could avoid the shock by jumping onto a platform at the rear of its enclosure during a warning signal, thereby preventing the shock from being given to itself or its partner. If the avoidance-escape rat failed to jump in time, so that shock occurred, it could still jump onto the platform to terminate the shock for itself and its partner. Thus the avoidance-escape rat could affect the occurrence and duration of shock by its responses, whereas its partner, called the yoked subject, received exactly the same shocks but was helpless: its responses had no effect at all on shock. The control rat, which received no shock, was simply allowed to explore its apparatus during the experiment. It is important to note that in all these experiments the three rats were assigned to their respec-



PREDICTABILITY EXPERIMENT showed that if a rat could predict when shocks would occur, it developed less gastric ulceration than a rat that received the same shocks unpredictably. One rat heard a "beep" before each shock, whereas for the other rat the "beep" occurred randomly with respect to the shock. A third rat, a control subject, heard the sound but received no shocks. Thus the only difference between the rats receiving shocks was psychological, and this psychological variable strongly affected the degree of ulceration.



STRESSFUL SITUATION (called "stressor") is created by the electric shock and the stimuli associated with the shock. When there is no warning-signal stimulus, the stressor condition extends through the entire time between shocks. When a warning of the shock is given, however, the stressor condition tends to be restricted to the warning signal.



COPING ATTEMPTS OR RESPONSES made by the rat in trying to escape from the stressor are related to ulcerogenic stress only if the responses fail to produce stimuli that are not associated with the stressor. If coping responses produce stimuli not associated with the stressor, the rat receives relevant feedback and ulcerogenic stress does not occur.

tive condition randomly just before the first trial.

The results showed that the yoked, helpless rats lost considerably more weight than the avoidance-escape rats. Over the several days of test sessions the yoked, helpless rats suffered an 80 percent reduction from "normal" weight gain (as measured in the control rats), whereas the reduction of weight gain in the avoidance-escape rats was only 30 percent. Again a psychological factor, in this case a difference in coping behavior, exerted a more powerful influence on stress responses than the occurrence of the physical stressor did.

A second experiment was immediately undertaken to test the generality of these findings, employing a different apparatus and measuring a different pathological response: gastric ulceration. In this case the avoidance-escape rat could avoid shock not by jumping onto a platform but by reaching through a hole in a small restraint cage to touch a panel mounted just outside. The avoidance-escape animal again had a yoked partner, which received the same shocks but was helpless, and a nonshock control. The rats were subjected to one continuous stress session lasting 21 hours, with the shocks—each preceded by a signal—scheduled to occur at the rate of one per minute. After the conclusion of the session all animals were sacrificed and their stomachs were examined for gastric lesions.

The effects of coping behavior were again found to be beneficial. The rats that were able to avoid and escape shock were found to show considerably less gastric ulceration than their yoked, helpless partners. And again the results pointed up how remarkable the effects of coping behavior could be. Whereas the stomach of the average avoidance-escape rat was found to have 1.6 millimeters of lesioned tissue, the average yoked rat had 4.5 millimeters of lesions, or roughly three times as much ulceration as the average avoidance-escape rat.

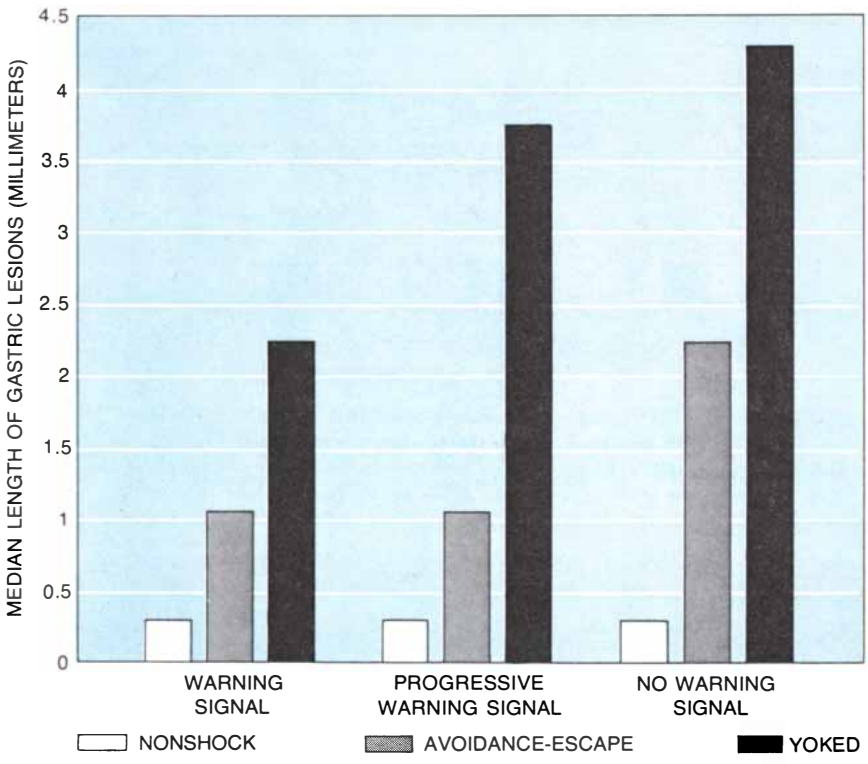
In both avoidance-escape experiments the shock was always preceded by a warning, so that the rats could predict when a shock was going to occur. Hence the avoidance-escape rat always had a signal to inform it when to respond. What would happen if there was no signal before the shock? Would the avoidance-escape rat again show less ulceration than a yoked subject?

To find out a large experiment was conducted in which three different warning-signal conditions were set up: no warning signal, a single uniform signal preceding the shock as in the earlier experiments, and a series of different signals that acted like a clock and therefore gave more information about when a shock would occur than the single uniform signal did. For these studies each rat was placed in a chamber with a large wheel [*see illustration on pages 104 and 105*]. If the avoidance-escape rat turned the wheel at the front of the apparatus, the shock was postponed for 200 seconds, or if shock had begun, it was immediately terminated and the next shock did not occur for 200 seconds. Thus the avoidance-escape conditions were exactly the same except for the difference in the warning signals. Each avoidance-escape rat had a yoked, helpless partner and both received exactly the same signals and shocks. A rat that never received a shock also was included in every case as a control subject.

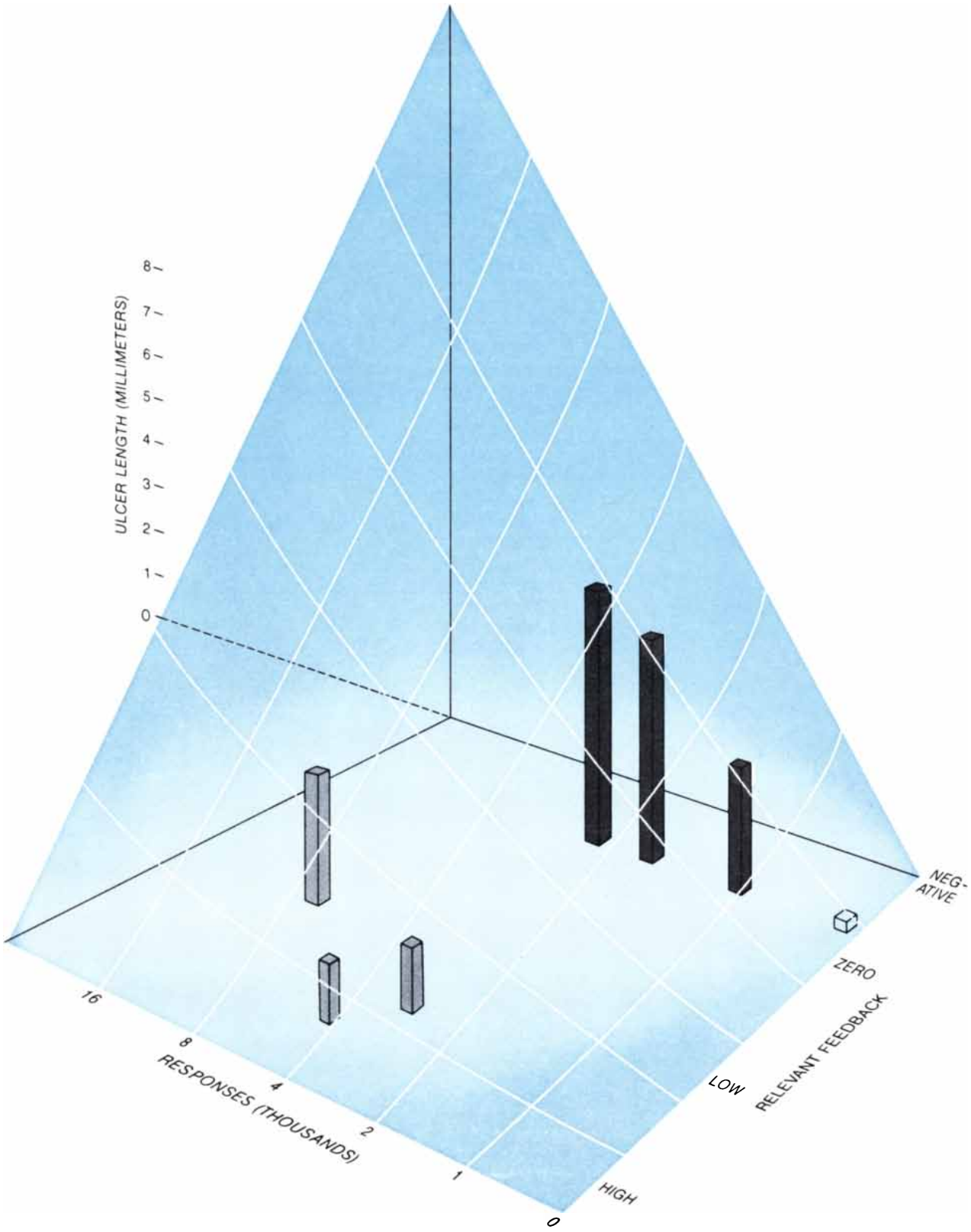
The results showed that regardless of the warning-signal condition avoidance-escape rats developed less gastric ulceration than yoked, helpless rats [*see illustration at left*]. Although the presence of a warning signal did reduce ulceration in both avoidance-escape and yoked groups, the avoidance-escape rats always developed less ulceration.

All the experimental findings on coping behavior that I have described up to this point have been opposite to the result found by Joseph V. Brady, Robert Porter and their colleagues in an experiment with monkeys [see "Ulcers in 'Executive' Monkeys," by Joseph V. Brady; SCIENTIFIC AMERICAN, October, 1958]. They reported that in four pairs of monkeys the animals that could avoid shock by pressing a lever developed severe gastrointestinal ulcers and died, whereas yoked animals that received the same shocks but could not perform the avoidance response survived with no apparent ill effects. Why were these results so markedly different?

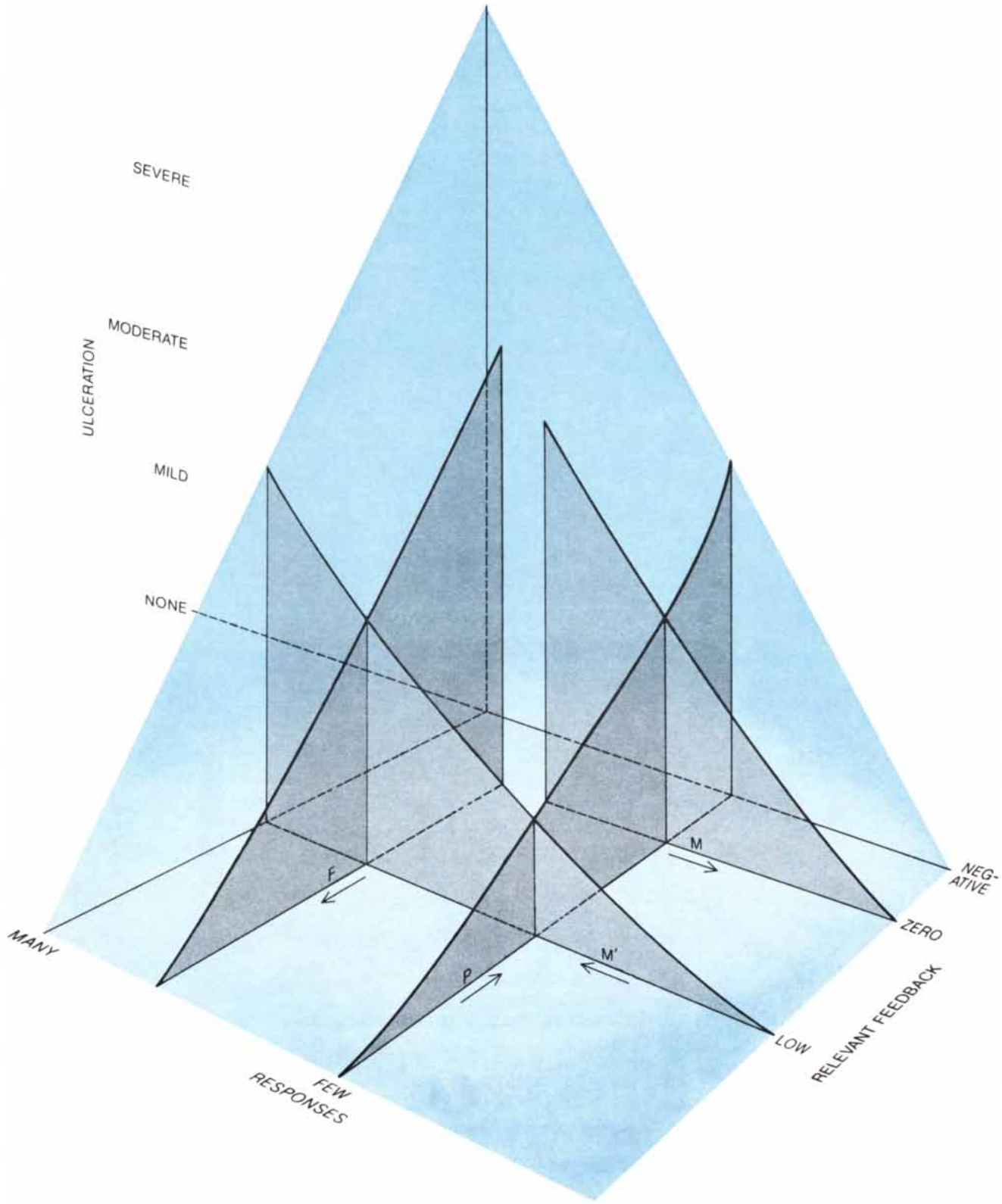
With careful study of the data from the 180 rats that were used in the coping-behavior and warning-signal experiment, it was possible to develop a theory that may explain how coping behavior affects gastric ulceration. This theory can account for the results I have



"HELPLESS" RATS develop more ulcers than their counterparts that can avoid or escape shock by performing a simple task, even though both rats have received exactly the same shocks. In all situations the avoidance-escape rat could terminate or postpone the shock by turning the wheel in front of it; its yoked partner received the same shocks but was unable to affect the shock sequence by its behavior. The control rat was never shocked. Regardless of the warning-signal conditions, rats that could do something to stop the shock developed less ulceration than their yoked, helpless mates. Ulceration was more extensive in both groups when there was no warning signal. The yoked, helpless rats unexpectedly developed almost as much ulceration with the progressive warning signal as with no warning signal.



RESPONSE RATE was found to be related to the amount of ulceration: the greater the number of responses, the more the ulceration. Moreover, increasing the amount of relevant feedback decreases ulceration. Combining these variables produces this three-dimensional graph. Here data from the illustration on the opposite page are replotted. The yoked, helpless rats given the progressive warning signal (*middle black bar*) made more responses than helpless rats given only a brief warning signal (*lowest black bar*). Helpless rats shocked without a warning signal made more responses than the other helpless rats and had the highest ulceration (*tallest black bar*). The single white bar represents the control rats that received no shocks in all three conditions, since their response rate and ulceration was nearly the same in all cases. (They developed some ulceration because they too were in a mildly stressful condition.) The avoidance-escape rats that received no warning signal made the greatest number of coping responses and had a high amount of ulceration (*tallest gray bar*). The avoidance-escape rats given the progressive warning signal (high feedback) made more responses (*middle gray bar*) than avoidance-escape rats that heard the brief warning signal before the shock (*right gray bar*).



PREDICTIVE MODEL of the relation between responses, feedback and ulceration is shown here. It predicts that the highest rate of ulceration will occur when the number of coping responses is very high and the feedback is very low or negative. The model can explain why the "executive" monkeys developed more ulcers than their "helpless" counterparts. The executive monkeys could postpone a shock by pressing a lever. Their response rate was quite high, but the amount of relevant feedback they received was low. Thus increased response rate along the low-feedback plane (*M′ in illustration*) led to increased ulceration in the executive monkeys. For the "helpless" monkeys, even though they received zero feedback (*Plane M*), their low response rate put them in the low-ulceration area. Similarly, if a high-relevant-feedback situation is changed into a negative-relevant-feedback situation (for example, a previously correct response suddenly begins to produce punishment), the model predicts a rapid increase in ulceration (*Plane P*) even at low levels of response. Finally, increasing feedback to a very high level should reduce ulceration to a very low amount (*Plane F*) even if the response rate is high. Results of experiments conducted to test these predictions are shown on page 112.

obtained and also can reconcile the seemingly contradictory results of the "executive monkey" experiment.

The theory states that stress ulceration is a function of two variables: the number of coping attempts or responses an animal makes, and the amount of appropriate, or "relevant," feedback these coping attempts produce. When an animal is presented with a stressor stimulus, the animal will make coping attempts or responses. The first proposition is simply that the more responses one observes, the greater is the ulcerogenic (ulcer-producing) stress. (Note that this does not say that the behavioral responses themselves cause ulceration, only that the amount of coping behavior and the amount of ulcerogenic stress tend to rise together.) If the responses, however, immediately produce appropriate feedback —that is, if the responses bring about stimuli that have no connection with the stressor—ulcerogenic stress will not occur. On the other hand, if the responses fail to produce such stimuli, then ulcerogenic stress will occur [*see bottom illustration on page 107*].

Perhaps the most important concept in this theory is that of feedback. The appropriate feedback is called relevant feedback. It consists of stimuli that are not associated with the stressful situation. Relevant feedback occurs when a response produces stimuli that differ from the stressor. The amount of relevant feedback produced depends on how different the stimulus situation becomes and how far removed these new stimuli are from any association with the stressor.

We can now specify how the two variables, responding and relevant feedback, are related to ulceration. Ulceration increases as the number of responses increases, and ulceration decreases as the amount of relevant feedback increases. Combining these two produces a function that forms a three-dimensional plane [*see illustration on page 109*]. From this model one can predict the amount of ulceration that is expected to occur in any stressful situation by specifying the number of coping attempts or responses the animal makes and the amount of relevant feedback these responses produce.

The model explains why animals able to perform effective coping responses usually develop fewer ulcers than helpless animals. Whenever a helpless animal makes a coping attempt, the response necessarily produces no relevant feedback because it has no effect on the stimuli of the animal's environment. Thus if helpless animals make an appreciable number of coping attempts, which many of them do, they will develop ulcers because of the lack of relevant feedback. Animals that have control in a stressful situation, however, do receive relevant feedback when they respond. In my experiments, for example, the avoidance-escape rats could terminate warning signals and shocks (thereby producing silence and the absence of shocks), so that their responses produced stimuli that were dissociated from the stressor. Hence animals in control of a stressor can usually make many responses and not develop ulcers because they normally receive a substantial amount of relevant feedback for responding.

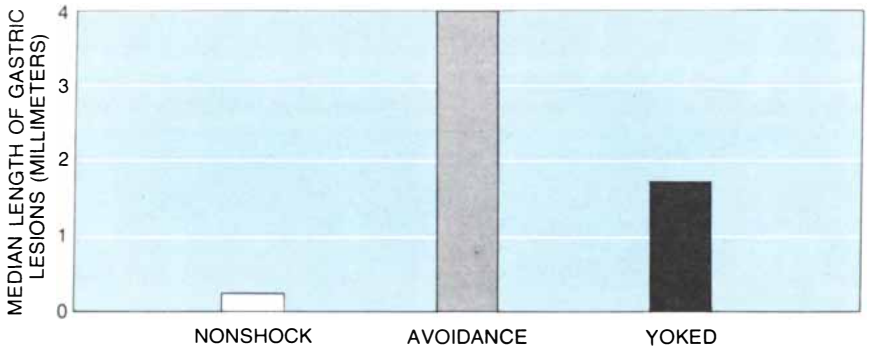
It is evident that, according to the theory, the effectiveness of coping behavior in preventing ulceration depends on the relevant feedback that coping responses produce; simply to have control over the stressor is in and of itself not beneficial. This means that conditions certainly can exist wherein an animal that has control will ulcerate severely. Specifically, in cases of low relevant feedback ulceration will be severe if the number of responses made is high [*see illustration on opposite page*].

I believe the foregoing statement tells us precisely why the executive monkeys died of severe gastrointestinal ulceration while performing an avoidance response. First of all, the responding of the monkeys was maintained at a very high rate in that experiment because they had to respond once every 20 seconds to avoid shock. In addition, the executives were actually selected for their high rate of responding. On the basis of a test before the experiment began, the monkey in each pair that responded at the higher rate was made the avoidance animal while its slower partner was assigned to the yoked position. Thus on the basis of their response rate the executive monkeys were more ulcer-prone from the beginning than their yoked partners. With regard to the relevant feedback for responding, the feedback for avoidance responding was quite low. There were no warning signals, and so the executives' rapid-fire responses could not turn off any external signals and therefore did not change the external-stimulus environment at all. As a result the relevant feedback came entirely from internal cues. Evidently this feedback was not sufficient to counteract the extremely high response condition, so that the executive animals developed ulcers and died. At the same time the yoked animals probably made very few responses or coping attempts because the shocks were few and far between, thanks to the high responding of the executives. It is no wonder, then, that the yoked animals in this case survived with no apparent ill effects.

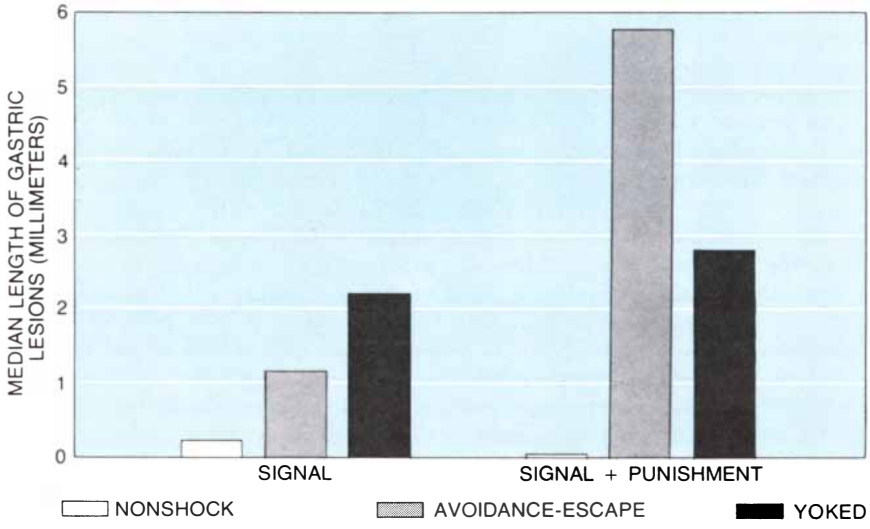
Further evidence in support of this model has emerged, both in an analysis of earlier experiments and in new direct tests with rats. Reviewing those experiments in which hyperactive avoidance-escape rats happened to be paired with low-responding yoked rats under conditions similar to those of the executive-monkey experiment, I found that high-responding avoidance-escape animals developed more ulceration than their helpless partners, which replicated the results of the monkey pairs [*see top illustration on next page*]. I then went on to test the model directly by examining the effects of very poor relevant feedback and excellent relevant feedback.

The first experiment examined the effects of very poor relevant feedback, which should, of course, produce severe ulceration. In this case avoidance-escape rats, having spent 24 hours in a normal avoidance situation with a warning signal, were given a brief pulse of shock every time they performed the correct response. Although the avoidance-escape rats had control over the stressor, their responses now produced the wrong kind of feedback: the stressor stimulus itself. The feedback in this condition is even worse than it is in the zero-relevant-feedback, or helplessness, condition. The results showed that even though these rats had control over the stressor, they developed severe gastric ulceration, in fact more ulceration than helpless animals receiving the same shocks [*see middle illustration on next page*].

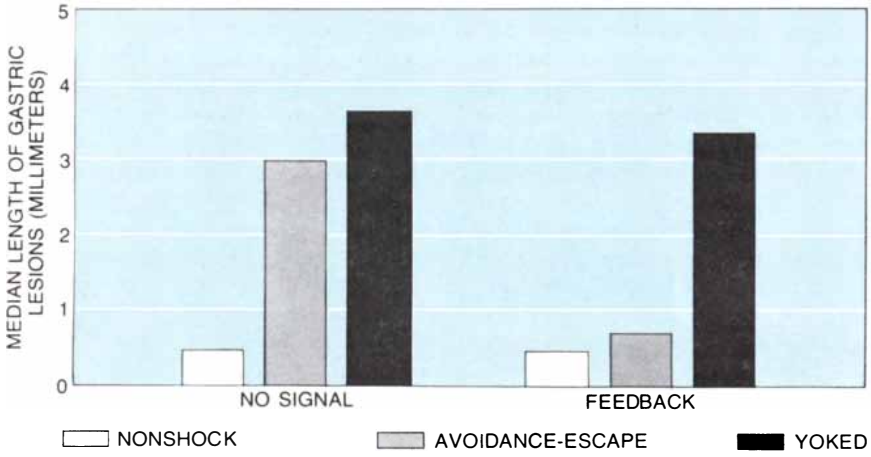
Having found that very poor feedback would cause severe ulceration, I conducted an experiment to determine if excellent feedback could reduce and possibly eliminate ulceration in a stressful situation. Initially the shock was administered without a warning signal, and under this condition the avoidance-escape rats will normally develop a considerable amount of gastric ulceration. Presumably ulceration occurs because the relevant feedback for responding is low, as in the case of the executive monkeys. Then a brief tone was added to the experiment. When the rat now performed its coping response, it not only postponed shock but also sounded the tone. Because the tone immediately followed the response, and the response postponed the shock, the tone was not associated with the shock. Thus the tone produced a change in the stimulus situa-



REPLICATION of the executive-monkey situation with data from experiments in which rats received unsignaled shocks offers support for the theoretical model of ulcerogenic stress. Matched pairs of high-responding avoidance rats and low-responding yoked rats were statistically selected and their ulcers measured. As the model predicts, the avoidance rats showed higher ulceration and the low-responding yoked rats had less ulceration.



NEGATIVE RELEVANT FEEDBACK produces severe ulceration even when the animal has control over the shock. In the warning-signal condition avoidance-escape rats learned to perform a response to avoid a shock whenever a tone sounded. In the punishment situation during the last half of the experiment the rats received a shock every time they performed the previously learned correct response. With this negative relevant feedback the avoidance-escape rats developed more ulcers than their yoked, helpless mates did.



EXCELLENT RELEVANT FEEDBACK following a correct response drastically reduces the amount of ulceration in rats. In the no-signal condition the avoidance-escape rat could postpone the shock by turning a wheel but there was no feedback other than the absence of shock. In the feedback condition a signal followed immediately after every correct response. The number of responses by avoidance-escape rats in both conditions was the same.

tion that constituted excellent relevant feedback. The result was striking. Although the rats in this situation received about as many shocks as the counterparts that were not given the tone feedback, they developed a small amount of ulceration; in fact, they developed only slightly more ulceration than controls receiving no shocks at all did [*see bottom illustration at left*].

Hence by manipulating the feedback consequences of responding, rats could be made to develop extensive gastric ulceration in an otherwise nonulcerogenic condition or could be protected almost completely from developing ulcers in a condition that was normally quite ulcerogenic. The fact that these results are consistent with the proposed model means that we are beginning to develop some idea of why the remarkable effects of psychological variables in stress situations occur.

It appears that the principles discovered in these animal experiments may operate in human situations as well. For example, Ronald Champion of the University of Sydney and James Geer and his associates at the State University of New York at Stony Brook have shown that if people are given inescapable shocks, the individuals who think they can terminate these shocks by clenching a fist or pressing a button show less emotional arousal as measured by electrical skin resistance than individuals who receive the same shocks and are also asked to clench a fist or press the button but are told that the shocks are inescapable. These findings can be explained using the model derived from the animal experiments, again emphasizing the role of relevant feedback. The people who thought they had control over the shock perceived their responses as producing the shock-free condition, that is, they saw their responses as producing relevant feedback. In contrast, the people who thought they were helpless necessarily perceived their responses as producing no relevant feedback. Thus for humans, as for rats, the same variables seem important in describing the effects of behavior in stress situations. On the other hand, the experiments with humans alert us to how important higher cognitive processes are in people, showing that verbal instructions and self-evaluation can determine feedback from behavior, which will subsequently affect bodily stress reactions.

Other stress responses have been measured in addition to gastric ulceration, for example the level of plasma corticosterone in the blood and the

amount of body weight the animals lost during the stress session. In many instances the results reflect those found with gastric ulcers, showing that ulceration may be only one manifestation of a more general systemic stress response. The correlation between measures is by no means perfect, so that it is evident that all physiological systems participating in stress reactions are not affected in the same way by a given stress condition. Certain systems in the body may be severely taxed by one set of conditions whereas other systems may be hardly affected at all, or actually may be benefited. For example, I have observed that heart weight tends to decrease in certain conditions, and this effect is seen more often in animals that are able to avoid and escape shock. We do not yet even know what this change indicates, but it suggests that certain conditions protecting one organ system, such as the gastrointestinal tract, might adversely affect another, such as the cardiovascular system.

Perhaps the most exciting biochemical system we have begun to study involves the catecholamines of the central nervous system. Eric A. Stone, Nell Harrell and I have studied changes in the level of norepinephrine in the brain. This substance is a suspected neurotransmitter that is thought to play a major role in mediating active, assertive responses, and several investigators have suggested that depletion of norepinephrine is instrumental in bringing about depression in humans. We found that animals able to avoid and escape shock showed an increase in the level of brain norepinephrine, whereas helpless animals, which received the same shocks, showed a decrease in norepinephrine. At the same time Martin Seligman, Steven Maier and Richard L. Solomon of the University of Pennsylvania have found that dogs given inescapable shocks will subsequently show signs of behavioral depression, but that dogs that are able to avoid and escape shocks do not show such depression. It may well be that the causal sequence leading from "helplessness" to behavioral depression depends on biochemical changes in the central nervous system, such as changes in brain norepinephrine. This would indicate that depressed behavior often can be perpetuated in a vicious circle: the inability to cope alters neural biochemistry, which further accentuates depression, increasing the inability to cope, which further alters neural biochemistry, and so on. We need to know more about this cycle and how to break it.

MATHEMATICAL GAMES

A miscellany of transcendental problems: simple to state but not at all easy to solve

by Martin Gardner

Pride in craftsmanship obligates the mathematicians of one generation to dispose of the unfinished business of their predecessors.

—E. T. BELL, *The Last Problem*

Two familiar irrational numbers are π (3.141...), which is the ratio of the circumference of a circle to its diameter, and e (2.718...), which is the base of natural logarithms. Each has a nonrepeating decimal fraction. Both π and e are also transcendental numbers, that is, numbers that are not algebraic. Specifically, a transcendental number is an irrational number that is not the root of an algebraic equation with rational coefficients. Is the sum of π and e transcendental? No mathematician knows if the sum is even irrational.

One might suppose that any two numbers with infinite, nonrepeating decimal fractions would necessarily have a sum with a nonrepeating (therefore irrational) decimal fraction. This is not the case. The difference between π and 7, for instance, is another transcendental. It is easy to compute. Represent 7 as 6.999..., then subtract π (3.14159...) to obtain 3.858407.... The sum of these two transcendentals obviously is 6.999..., or 7.

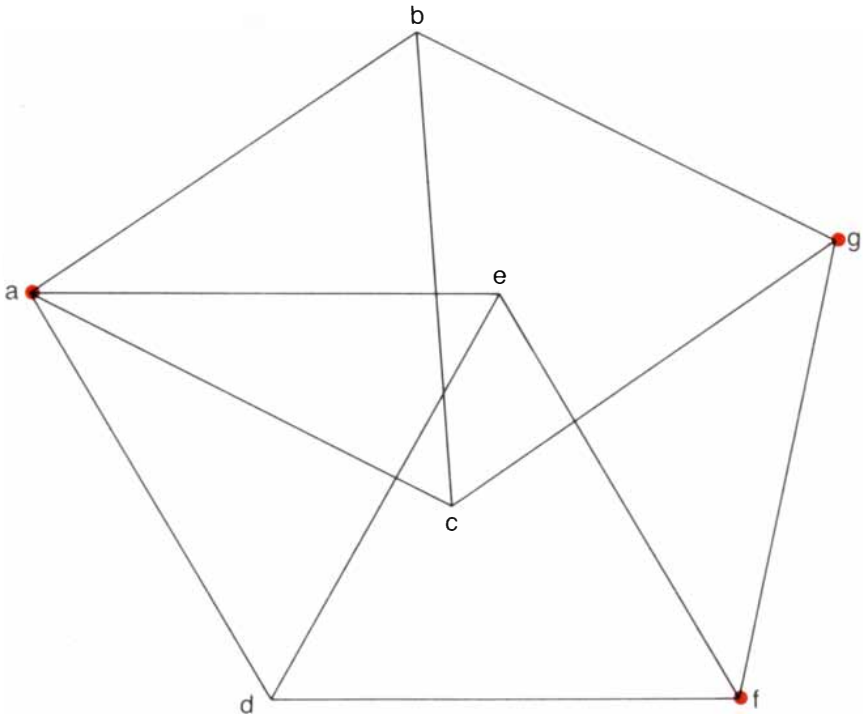
It seems unlikely, but until someone proves otherwise π and e could be related by a curious unknown formula that would give their sum a repeating (rational) decimal fraction with a very long period, say one of several million digits. It also is not known if πe, π^{π}, e^{e} or π^{e} are irrational. It *has* been shown, however, that e^{π} is transcendental, and last year David A. Brubaker proved that at least one of the two numbers, πe and $(\pi + e)$, is transcendental. The unanswered questions about π and e are among hundreds of problems that are ridiculously simple to state but so difficult and deep that long-lasting fame awaits the first person to solve them.

It is not easy to distinguish significant unsolved problems from trivial ones. In *A Mathematician's Apology* G. H. Hardy characterized a significant problem as being one connected to such a large complex of other mathematical ideas that when it is solved, it leads to important advances in mathematics and perhaps in science as well. An example of an essentially trivial but extremely difficult question is: If two people play the best possible checker game, will it end in a draw, a victory for the player who makes the first move or a victory for the player who makes the second move? A computer, given enough time, will probably work out the answer one day. When it does, the solution is unlikely to lead to any breakthroughs in mathematics or science. On the other hand, settling the four-color map conjecture or Fermat's last theorem would open all kinds of barred doors. (Please do not send me proofs. I am incapable of spotting flaws and always return them unread.)

There are dozens of unsolved map-coloring problems that, although they may not be as profound as the four-color theorem, are by no means trivial. Here is a notorious one given in C. Stanley Ogilvy's new revision of *Tomorrow's Math* (Oxford, 1972), a splendid collection of unsolved problems for amateurs. What is the minimum number of colors needed to color the plane in such a way that any pair of points a unit distance apart are in regions of different colors? The question was first raised 20 years ago by Paul Erdös, a prolific inventor of problems.

That such a map must have at least four colors was cleverly established by Leo Moser with the diagram in the illustration at the left. Each edge of this graph has a length of one unit. Imagine that the graph is placed anywhere on a plane in which the problem is solved with only three colors. If vertex *a* is on red, say, then *b* and *c* must be on the other two colors, and *g* also must be red. Similarly, *d* and *e* must be on the other two colors and *f* must be red. Now, however, we have contradicted our assumption, because *f* and *g*, which are a unit distance apart, are both on red. At least four colors are therefore necessary.

Ogilvy's book has a proof that seven



Leo Moser's graph for proving four-color necessity

different colors are enough [*see illustration at right*]. The numbers give a repeating color pattern for a hexagonal tessellation of the plane, each hexagon slightly less than one unit from corner to opposite corner. The gap remaining to be closed is a big one. Do such maps of four, five or six colors exist? No one as yet knows.

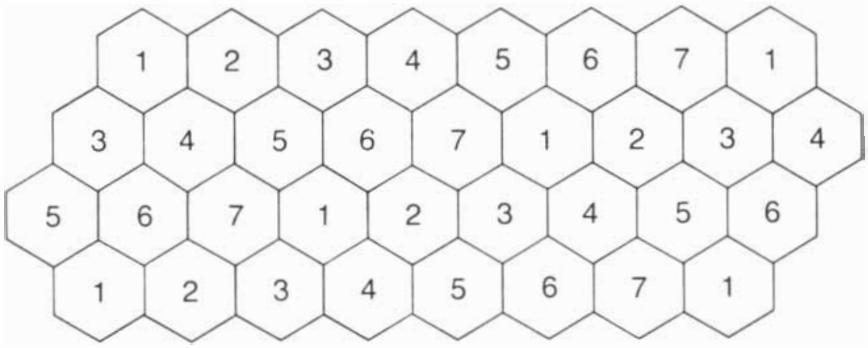
There is an unusual class of unsolved arithmetic problems that, to use a computer-science term, we can call "looping" problems. A series of integers is generated according to a rule. One then asks if the series always enters one or more loops in which a finite set of integers keep repeating cyclically. For example, start with any positive integer. Halve it if it is even; triple it and add 1 if it is odd. Keep repeating this procedure until the series loops in the cycle 2, 1, 4, 2, 1, 4, ... (Sample: 3, 10, 5, 16, 8, 4, 2, 1, ...) But does the series always enter the 2, 1, 4 loop? No one has proved that it does, nor has a counterexample been found.

Since Ogilvy revised his book a group of workers in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology have computer-tested all positive integers to 60,000,000 without finding an exception. They also discovered that if the rule $3n + 1$ (for odd integers) is replaced by $3n - 1$, the result, in absolute values, is the same as starting with a negative integer and following the old rules. In this case all negative integers to $-100{,}000{,}000$ were found to enter one of three loops with the following absolute values:

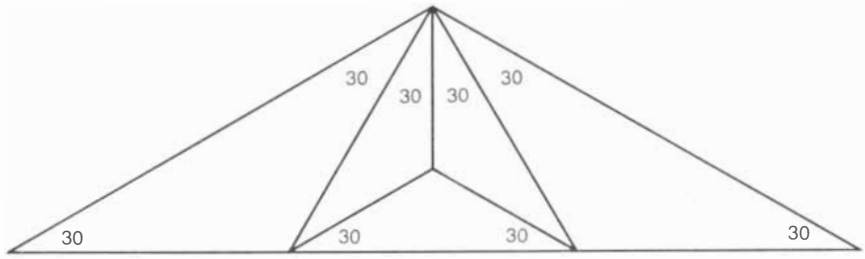
1. 2, 1, 2, 1, ...
2. 5, 14, 7, 20, 10, 5, ...
3. 17, 50, 25, 74, 37, 110, 55, 164, 82, 41, 122, 61, 182, 91, 272, 136, 68, 34, 17, ...

Michael Beeler, William Gosper and Rich Schroeppel give these results in HAKMEM (short for "Hacker's Memo"), Memo 239, Artificial Intelligence Laboratory, M.I.T., 1972, page 64. No one has yet come up with good ideas about how to establish the general case for all nonzero integers. (Zero, of course, is already in a 0, 0, 0, ... loop.) No one knows if there are other loops, or if there are integers that generate a nonlooping series of numbers that diverge to infinity.

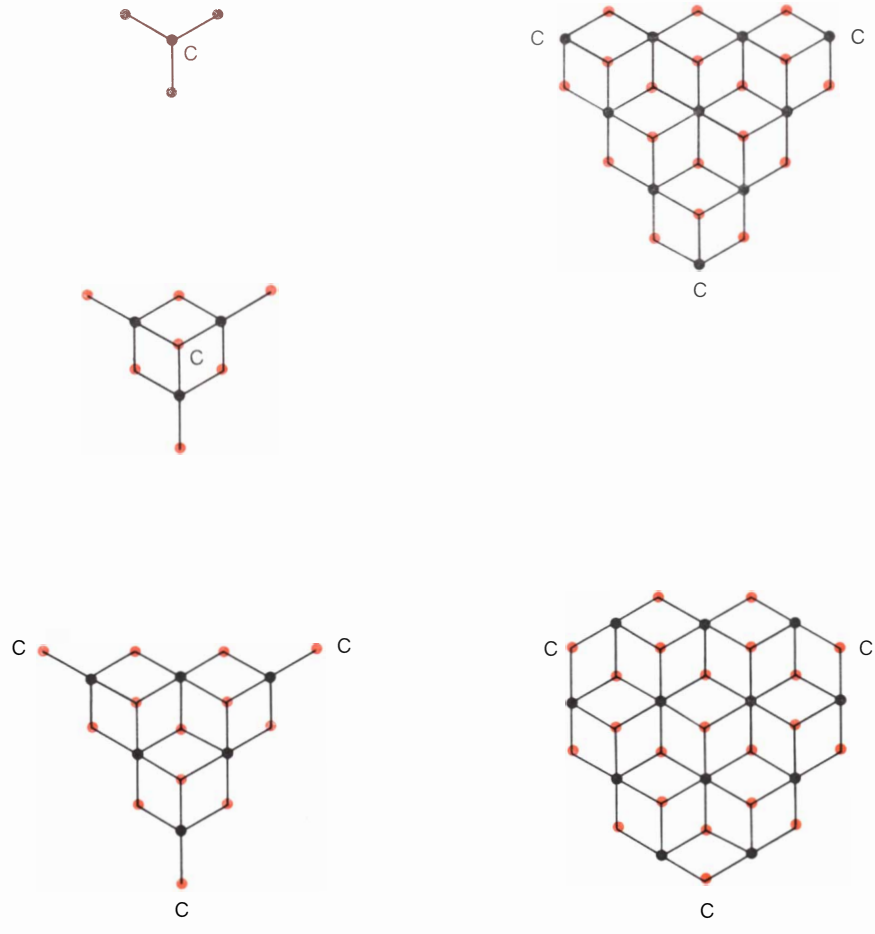
Gosper and Schroeppel, incidentally, proved an amusing loop conjecture involving English names for numbers (HAKMEM, page 64). Spell out the name of any number. It need not be rational



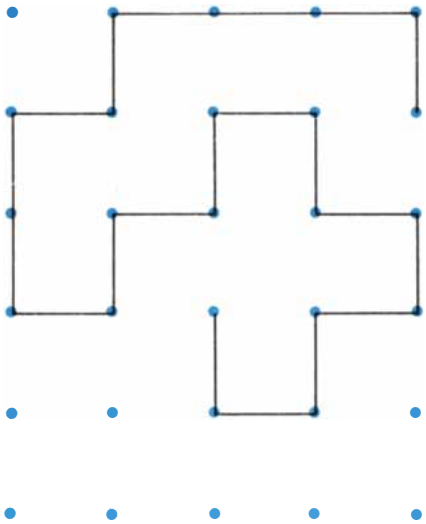
Proof of seven-color sufficiency



A unique dissection



Graphs for Stanislaw M. Ulam's knot game



The game of Slither

or even real. Counting numbers must be named directly, and not by such circumlocutions as "twelve plus one" or "twenty minus five," and so on. Replace the name by the number of digits in the name and keep repeating the procedure. Example: THE CUBE ROOT OF PI, FIFTEEN, SEVEN, FIVE, FOUR, FOUR, FOUR, ... The series always, and quickly, loops at FOUR.

In explaining a recent triangle-dissection problem, Ogilvy wrote that it might "have a solution before this book appears." He was right. It had been known that any triangle can be cut into four triangles similar to itself, or into n triangles similar to itself when n is 6 or more. If n is 2 or 3, only a right triangle can be properly cut. If n is 5, a right triangle can be dissected into five triangles similar to itself, but for nonright triangles the conditions for dissection were unknown when R. W. Freese, Ann K. Miller and Zalman Usiskin wrote their article "Can Every Triangle be Divided into n Triangles Similar to It?" (*The American Mathematical Monthly*, Volume 77, October, 1970, pages 867–869).

Recently it has been independently proved by several people that when n is 5 and the triangle has no right angle, it can be cut into five triangles similar to itself if and only if one angle is 120 degrees and the others are each 30 degrees [*see middle illustration on preceding page*]. This unique dissection is given in "Partitioning a Triangle into 5 Triangles Similar to It" (*Mathematics Magazine*, Volume 45, January, 1972, pages 37–42). Still open are questions such as: Which triangles can be cut into n similar triangles *not* similar to themselves? For what values of n can a quadrilateral be cut into n quadrilaterals similar to one another and/or to itself?

NODES	LENGTH	DIVISIONS
3	3	1, 2
4	6	1, 3, 2
5	11	1, 3, 5, 2
		2, 5, 1, 3
6	17	1, 3, 6, 2, 5
		1, 3, 6, 5, 2
		1, 7, 3, 2, 4
		1, 7, 4, 2, 3
7	25	1, 3, 6, 8, 5, 2
		1, 6, 4, 9, 3, 2
		1, 10, 5, 3, 4, 2
		2, 1, 7, 6, 5, 4
		2, 5, 6, 8, 1, 3
8	34	1, 3, 5, 6, 7, 10, 2
9	44	1, 4, 7, 13, 2, 8, 6, 3
10	55	1, 5, 4, 13, 3, 8, 7, 12, 2
11	72	1, 3, 9, 15, 5, 14, 7, 10, 6, 2
		1, 8, 10, 5, 7, 21, 4, 2, 11, 3

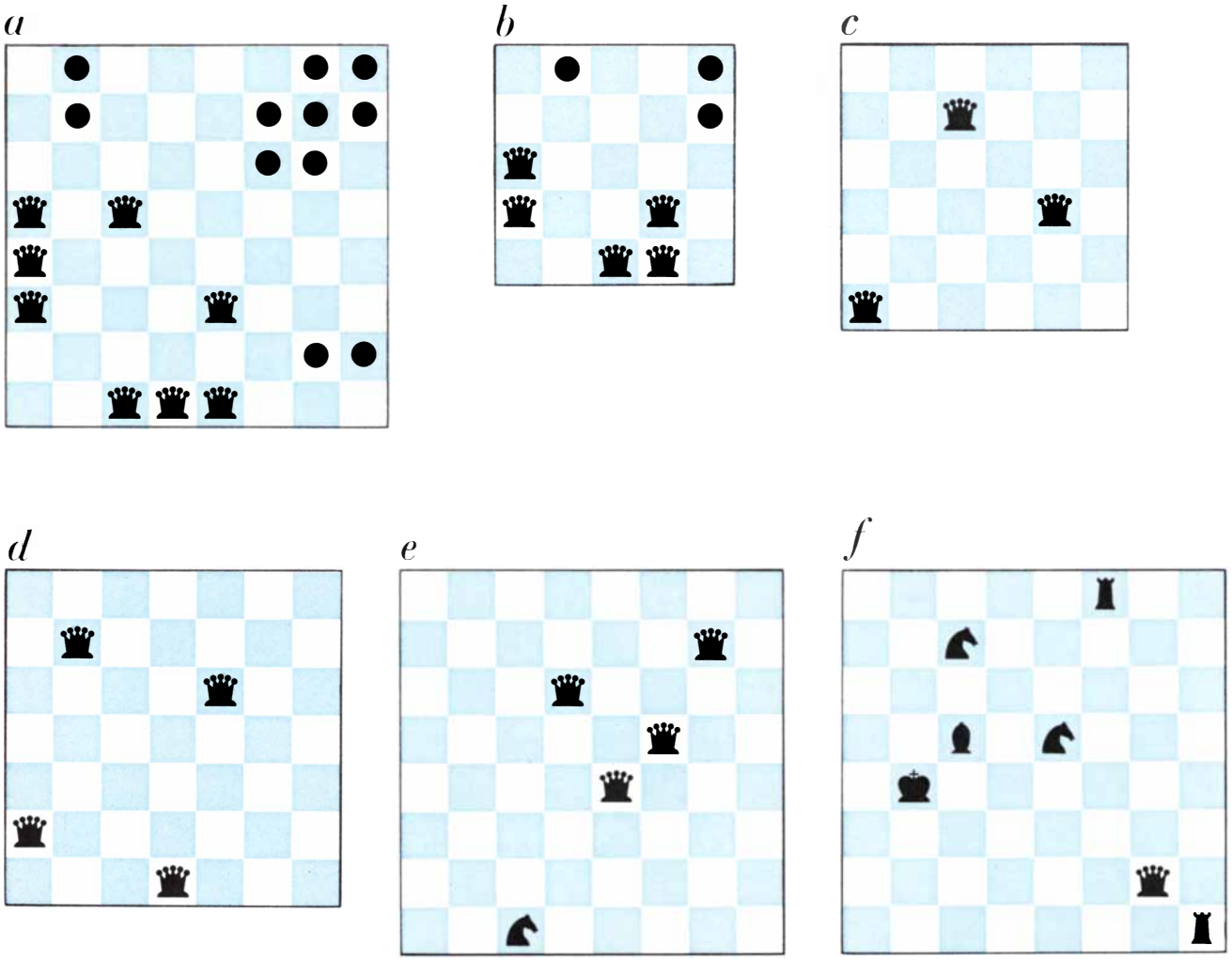
Minimum-length Golomb rulers

In 1960 Stanislaw M. Ulam, another virtuoso puzzle maker, published a fine collection of advanced unsolved problems, most of them original. The book was reprinted in 1964 as a paperback, *Problems in Modern Mathematics* (John Wiley & Sons). One of Ulam's topological-game problems seemed as uncrackable as it was curious. Imagine a cube divided into a lattice of unit cubes, like a three-dimensional checkerboard. Players take turns marking a unit edge of the lattice. The first player marks any edge. Thereafter each marked edge must join the previously marked edge. One end of the path remains fixed as the other end grows one unit in length with each move, as though a bug were crawling along the lattice lines and leaving a trail. Since the lattice is finite, the path must eventually intersect itself to form a closed-space curve. One of the players wins if the curve is knotted. The other player wins if there is no knot. Who wins when the game is played rationally?

John Horton Conway, a University of Cambridge mathematician now on leave at the California Institute of Technology, found an ingenious proof that the "no knot" player can always win regardless of whether he goes first or second. Assume that the game is played on a three-by-three-by-three cube (a lattice of 27 unit cubes). This is the smallest cube on which the path can knot. The following no-knot strategy extends readily to all cubes of higher orders.

Through each lattice point there are various planes that are perpendicular to a body diagonal (a diagonal joining diametrically opposite corners) of the large cube. We shall call such a plane a primary plane, or P plane. If the P plane goes through a corner of the large cube, there may be only one adjacent plane parallel to it and passing through points adjacent to points on the P plane; otherwise there will be two such adjacent planes, one on each side of the P plane. We call these A planes.

Imagine all lattice points on A planes —call them A points—projected on the P plane, together with all edges joining A points to P points. This puts on the P plane a graph equivalent to one of the five shown in the bottom illustration on the preceding page. On each graph black



Answers to last month's chess tasks

vertexes are lattice points originally on the *P* plane. Colored vertexes are *A* points projected from *A* planes. The *C*'s mark corners of the large cube.

Note that the three graphs on the left have loose ends. Those on the right do not. Conway has shown (his unpublished proof is not difficult) that, given any lattice point, one can always find passing through it a *P* plane on which the graph has no loose ends.

If your opponent goes first, you should choose a *P* plane through either end of the marked edge that has a no-loose-ends graph. One end of the path will be on an *A* plane. Play so that the path returns to the *P* plane, that is, join the colored end of the path to a black vertex. Each succeeding move by your opponent must take the path off the *P* plane (to a colored spot). Your strategy is always to return the path to the *P* plane by extending it to a black spot. Because the graph has no loose ends, and because its colored and uncolored vertexes alternate, you can always do so. It is obvious that when the path first closes, it will have to be unknotted.

If you go first, mark any edge. After your opponent has moved choose a *P* plane that goes through the path's middle vertex and on which the graph has no loose ends. Your no-knot strategy is the same as before. Play so that the path always returns to the *P* plane; in other words, always extend the path to a black vertex. The path cannot be knotted when it first intersects itself.

"I think it was obvious from the start," Conway writes in a letter, "that the no-knot player had the best of it. *He* only had to make the path close, whereas the other player really had to *do things*."

Conway's strategy does not apply to noncubical "brick" lattices (because finding a no-loose-ends graph is not always possible) or to cubical games in which the no-knot player goes first and moves are allowed at all times at either end of the growing path. In both cases, so far as I know, winning strategies remain unknown.

Three-dimensional lattices are awkward "boards" for actual play, but closely related topological games on planar lattices make excellent pencil-and-paper contests. David L. Silverman, whose book *Your Move* includes several such games, is responsible for the latest fad among Los Angeles puzzlers: an unpublished, unsolved game that Silverman calls Slither. Its five-by-six-point lattice is just large enough to have resisted all efforts to determine which player has the win [*see top illustration on opposite page*]. In a tabulation of several hundred games the wins were about equally divided between first and second players. The rules are simple. Opponents take turns marking an orthogonal unit segment. The segments must form a continuous path but may be added to either end of the preceding path. The player forced to close the path is the loser. (If the first to close it wins, it is a duller

game, although even *that* version is unsolved.) The illustration shows a typical position in which the next play must be a losing one. Perhaps a reader will discover a winning strategy for either version (or both versions) of Slither.

Hallard T. Croft, a colleague of Conway's at Cambridge, periodically sends lists of new unsolved problems to his friends. A few years ago one of Croft's problems asked if there existed a finite set of points on the plane such that the perpendicular bisector of the line segment joining any two points would always pass through at least two other points of the set. The problem was solved by Leroy M. Kelly, a mathematician at Michigan State University. Although the problem cannot be called significant, Kelly's solutions, using only eight points, is so elegant that I give it as an exercise to be answered here next month.

In March readers were asked to improve on the rows of "Golomb's triangle," each row giving the shortest-known rulers of n marks (including end points) such that every distance between a pair of marks is a distinct integer. Walter Penney of Greenbelt, Md., was the first to lower the eight-mark ruler to length 34. The same ruler was also found by hand by Daniel A. Lynch of Wildwood, N.J.

William Mixon of the University of Chicago was the first to make an exhaustive computer search for all minimum-length rulers through 11 marks. His results show that rulers of eight, nine and 10 marks are unique, except of course for reversals [*see bottom illustration on page 116*]. These results were completely confirmed by Ashok Kumar Chandra's computer program at Stanford University and partly confirmed by the programs of Paul Steier, James R. Van Zandt, Edward Schonberg and others. Working by hand, Sheldon B. Akers found the nine-mark ruler, and Wolfgang Harries, also working by hand, found all but one of the rulers with six and seven marks.

The 10-mark ruler and one 11-mark ruler had been found earlier by John P. Robinson of the University of Iowa with a nonexhaustive computer search made in connection with work on his 1966 doctorate on error-correcting codes. His best results, for rulers through 24 marks, are given in "A Class of Binary Recurrent Codes with Limited Error Propagation," by Robinson and Arthur J. Bernstein, in *IEEE Transactions on Information Theory* (Volume IT-13, Number 1, January, 1967, pages 106–113).

R. C. Ashenfelter of the Bell Telephone Laboratories was the first to gracefully label the dodecahedron. Chandra devised a computer program that made an exhaustive search for the icosahedron and produced five fundamentally different labelings. A partial search for the dodecahedron yielded a large number of graceful labelings. This settles affirmatively Golomb's conjecture that the skeletons of all five Platonic solids are graceful graphs.

Last month's chess tasks have the following answers:

1. Sam Loyd's three-move mate, all white men in starting position and a lone black king on Black's KR5:

1. P–Q4	1. K–R4
2. Q–Q3	2. K moves
3. Q–KR3 (mate)	
or	
1. P–Q4	1. K–N5
2. P–K4 (ch)	2. K moves
3. P–KN3 (mate)	

2. One of six known ways to place eight queens so that 11 vacant cells are unattacked is shown in illustration *a* on the preceding page. The unchecked squares are indicated by dots.

3. There is only one basic way to place five queens on an order-5 board so that three vacant cells are unattacked [*see illustration b on preceding page*]. Mannis Charosh has suggested that the best systematic search procedure for proving uniqueness is to explore the equivalent problem of placing three queens so that five cells are unchecked, taking advantage of the board's symmetries to shorten the search. I shall report later on any improvements by readers of the maximums given last month for boards of order 6 through order 12.

4. The only basic way to place three queens on an order-6 board so that all vacant cells are checked is shown in illustration *c*.

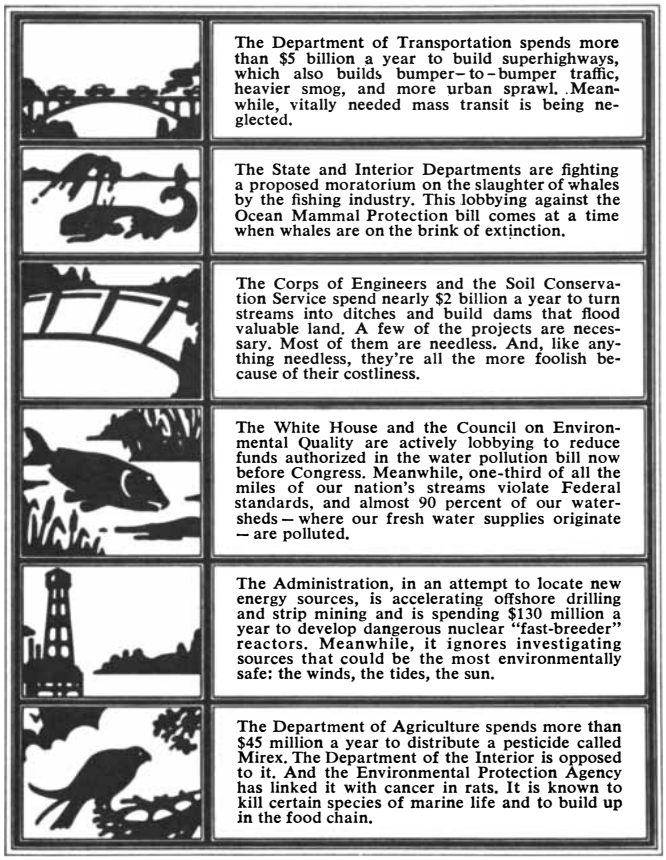
5. The only basic way to put four queens on an order-7 board so that all vacant cells are checked and no queen attacks another queen is shown in illustration *d*.

6. The only known basic way to place four queens and one knight so that all vacant cells are attacked is shown in illustration *e*.

7. Illustration *f* shows one way to place seven of the eight pieces of one color so that all vacant cells are in check. The positions of the rooks and queen can be given trivial variations.

THE GREAT MISTAKES CONTEST

Wrecking the environment seems to be a hot competition among most government agencies. The blunders are usually made on your tax money. Here are a few expensive examples:



The Department of Transportation spends more than $5 billion a year to build superhighways, which also builds bumper-to-bumper traffic, heavier smog, and more urban sprawl. Meanwhile, vitally needed mass transit is being neglected.

The State and Interior Departments are fighting a proposed moratorium on the slaughter of whales by the fishing industry. This lobbying against the Ocean Mammal Protection bill comes at a time when whales are on the brink of extinction.

The Corps of Engineers and the Soil Conservation Service spend nearly $2 billion a year to turn streams into ditches and build dams that flood valuable land. A few of the projects are necessary. Most of them are needless. And, like anything needless, they're all the more foolish because of their costliness.

The White House and the Council on Environmental Quality are actively lobbying to reduce funds authorized in the water pollution bill now before Congress. Meanwhile, one-third of all the miles of our nation's streams violate Federal standards, and almost 90 percent of our watersheds – where our fresh water supplies originate – are polluted.

The Administration, in an attempt to locate new energy sources, is accelerating offshore drilling and strip mining and is spending $130 million a year to develop dangerous nuclear "fast-breeder" reactors. Meanwhile, it ignores investigating sources that could be the most environmentally safe: the winds, the tides, the sun.

The Department of Agriculture spends more than $45 million a year to distribute a pesticide called Mirex. The Department of the Interior is opposed to it. And the Environmental Protection Agency has linked it with cancer in rats. It is known to kill certain species of marine life and to build up in the food chain.

It's time to introduce the prime cause of environment problems. Meet the government. Some of whose agencies allow billions of dollars' worth of mistakes and don't stop the mistakes once they've been started.

You can help put an end to the blundering.

Friends of the Earth has accomplished the near-impossible before. We intend to keep doing it. For $15, you can join with us. And help us keep more mistakes from getting out of hand.

Friends of the Earth knows how to stop the mistakes before they become earth-shattering blunders.

Besides organizing and directing the campaign that stopped the SST, we've also helped lead the fight to relocate the jetport that could have killed Everglades National Park. We lobbied for the Floridians who stopped the Cross-Florida Barge Canal. And are joining with other groups to help delay the Trans-Alaska Pipeline until a more reasonable alternative can be found.

We don't take full credit for blunder-cutting. We've worked closely with every one of the most effective environment groups in this country. And we go farther. Working in Germany, France, Holland, Sweden and the United Kingdom, for example. As international conservation activists we've learned that the environment has no borders.

Own a share of the FOE environmental lobby in Washington.

Your membership will help our four paid lobbyists and many volunteers in their combat with the many hundreds of industry lobbyists fighting to protect their power to pollute.

We do it first by providing the necessary research and information for the lobbying staffs. When a Congressman calls to find out just how the SST pollutes the air and the eardrums, FOE has the facts and arguments necessary to defeat it.

Next, we alert his constituents and bring some of them to Washington. The more influential they are, the better. They tell him to his face they don't want the SST.

Business writes off its lobbying expenses. We can't. That's why we need you.

It's hard to raise money for a non-deductible organization. But it's only through your contribution that we can keep our four registered lobbyists going strong. Your contribution and our influence stand a strong chance of stopping another blunder.

Left on its own, the government isn't going to do it for you.

Help call off The Great Mistakes Contest.

Spread this ad around. Copy it for your club magazine. Your company house organ. Your dentist's wall or your favorite bistro bulletin board. Get stores to use it as a bill stuffer. Use your influence with the local newspapers, television and radio stations. Use your voice and your imagination.

But most of all, help us to continue our existence by your contribution. Because, since The Great Mistakes Contest shows no sign of ending, you and Friends of the Earth need all the help we can get.

FRIENDS OF THE EARTH
529 Commercial Street, Box 260H
San Francisco, California 94111

$_____Enclosed is my contribution for lobbying to stop the blunders. Please enroll me as: ☐ Regular Member – $15 yr. ☐ Supporting Member – $25 yr. ☐ Contributing Member – $50 yr. ☐ Sustaining Member – $250 yr. ☐ Life Member – $1000 or more ☐ Student Member – $7.50. ☐ Included is $5 additional for Spouse Membership.

Contributions to Friends of the Earth are not deductible.

Supporting Members will receive six free books from our series of paperbacks. Contributing, Sustaining and Life Members will receive the six paperbacks plus any volume of their choosing from "The Earth's Wild Places" exhibit book series. $3.00 of each membership is for a subscription to the monthly newspaper "Not Man Apart."

If you prefer to make a tax deductible contribution:

JOHN MUIR INSTITUTE FOR ENVIRONMENTAL STUDIES
(non-lobbying research affiliate of FOE)
c/o Friends of the Earth (above address)

Enclosed is my contribution of $_____.

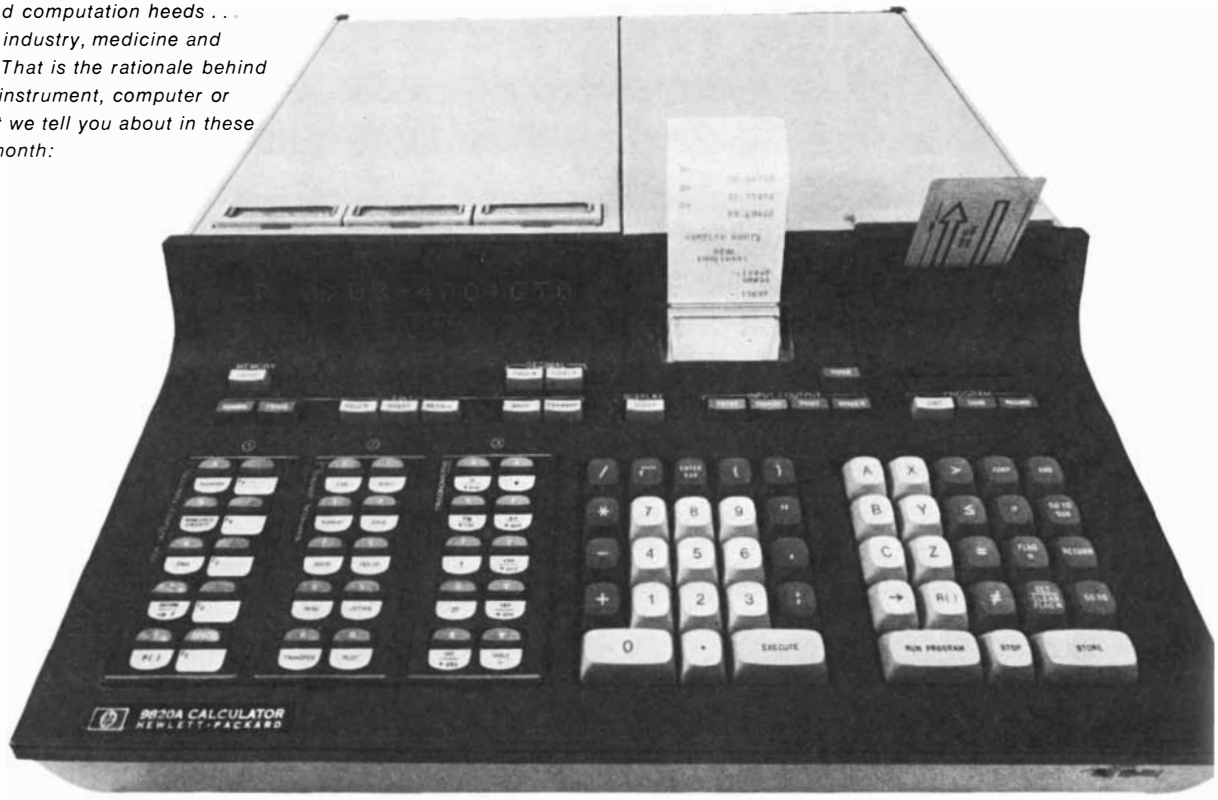
Name (Please Print)_____

Address_____

City, State, Zip_____

Some things are changing for the better.

Many people know us as an instrument manufacturer: we make more than 2,000 products for measurement, test and analysis. Others know us as a computer company: more than 10,000 own our programmable calculators and computers. We prefer to think that our business is to serve measurement, analysis and computation heeds . . . in science, industry, medicine and education. That is the rationale behind every new instrument, computer or system that we tell you about in these ads. This month:



Powerful new programmable calculator converses in simple algebraic language.

Hardly a month goes by that doesn't signal the introduction of a new calculator with more power, more memory, more output flexibility. But are things really changing for the better when the improvements so complicate the use of a calculator that one must practically become a computer programmer before he can harness its power?

This is precisely where HP's new Model 20 makes its most significant contribution. It is easier to use, by far, than any other calculator. Its language is the most simple: algebra, the kind you learned in high school. Not only does it 'understand' algebra, it also 'speaks' it, using the same numbers, letters and symbols that you do.

You enter an equation just as you write it on paper including implied multiplication and nested parentheses. The Model 20 displays your entry for verification the same way you wrote it. For example:

```
(-B+√(BB-4AC))/2A
```

It tells you what to do next (in the program mode):

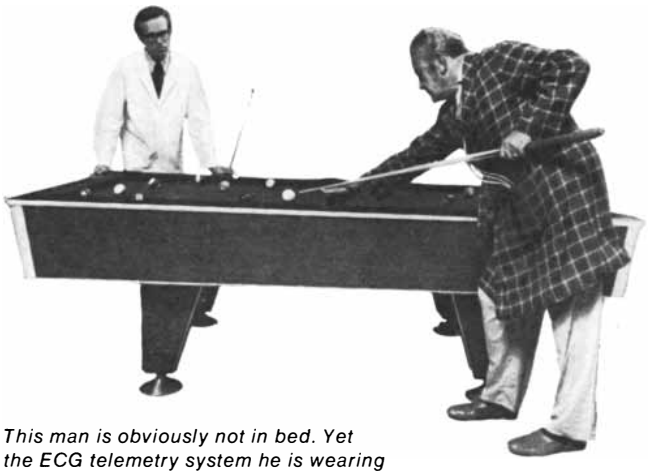
```
ENTER A
```

Then you enter the values on the keyboard (for A, B and C in this example) and press a key to execute. The calculator immediately displays and prints the solution to ten significant digits, along with its English label if you desire:

```
REAL ROOTS
```

If you make a mistake, the Model 20 tells you so and identifies precisely what the error is . . . lets you correct it without redoing the entire line, let alone the entire program . . . and automatically adjusts program storage to occupy the least possible memory.

In addition to its conversational ability, the Model 20 changes things for the better not only through more power and more memory but through a hardworking line of Series 9800 Peripherals: X-Y Plotter, Typewriter and Card Reader, to name a few. Model 20 costs $5,475.

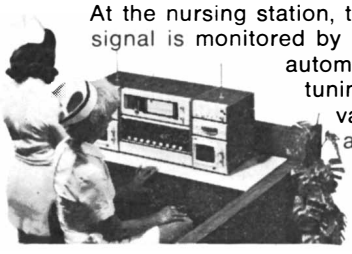


This man is obviously not in bed. Yet the ECG telemetry system he is wearing enables nurses at a central monitoring station to keep close watch on his heart action.

Freedom with protection for the post-coronary patient.

Once the coronary patient is released from the intensive care unit, his recovery can often be aided by freedom to move about and mild exercise . . . provided his ECG can be continuously monitored.

With the new HP ECG Telemetry System, the post-coronary patient can be ambulatory. Wherever he goes, his heart action is transmitted to a receiver at the nursing station where it can be continuously observed. The transmitter is small enough to be carried comfortably in a bathrobe pocket, has a strong enough signal to reach the nursing station from 200 feet even through several masonry walls, and is rugged enough to operate reliably even if dropped.



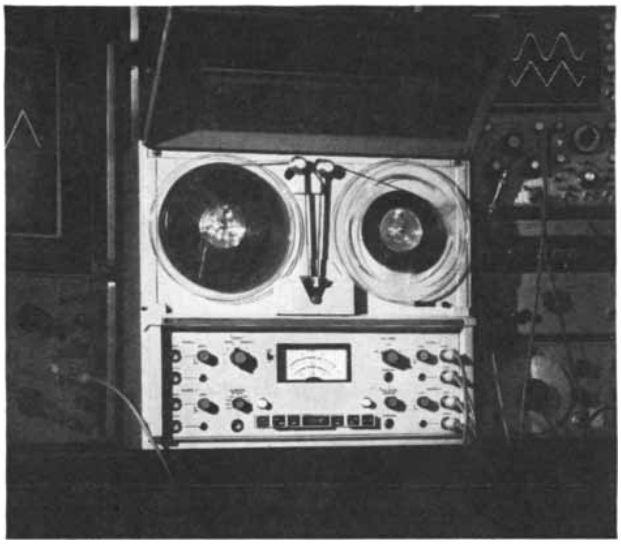
At the nursing station, the patient's ECG signal is monitored by a receiver that operates automatically, never requires tuning and accepts only valid signals, minimizing artifacts from patient motion. An automatic warning light alerts the nurse of such inoperative conditions as: patient out of range, dislodged electrode, low battery power. It is completely compatible with HP patient monitoring systems. Because it doesn't require new wiring, the ECG Telemetry System is easily introduced into existing facilities. Price is $1,800 for each patient unit. Write for our new illustrated brochure.

Instrumentation quality tape recording at a bargain price.

Most scientists would use portable instrumentation tape recorders for analog recording if only they performed as well as the big expensive laboratory machines. Unfortunately, their small size usually meant reduced performance.

Then came the HP 3960. Truly portable in size (50 pounds) and low in price ($4,270 for a fully-equipped four-channel instrument), the 3960 actually outperforms most laboratory machines costing five times more.

If this sounds too good to be true, listen to some of the 3960's capabilities. At 15/16 ips, its FM signal-to-noise ratio of better than 200:1 lets you play back signals that would be buried in noise (ECG's for example) on many lab machines as well as on any other portable.



The 3960 lets you mix and interchange four FM direct record/reproduce channels at will. You have a choice of three electrically-switched speeds, for a time-base expansion of 16:1 or 10:1 . . . without signal degradation. Tape drive is bidirectional so that you don't have to rewind either to continue recording or to play back.

Built-in facilities let you calibrate the 3960's FM electronics without external equipment. And an integral peak-reading meter lets you optimize record level without using a scope. Options include a 5 to 30 foot loop adaptor, an interrupting voice channel, and an inverter for 12 or 28 VDC . . . all integrally mounted.

Write for Application Note 89, a tape recording handbook useful to scientists interested in tape recording techniques for vibration and test analysis, research and clinical medicine, acoustics, oceanography and other environmentally difficult research projects.

For complete information, write Hewlett-Packard, 1505 Page Mill Road, Palo Alto, California 94304. In Europe: 1217 Meyrin-Geneva, Switzerland; Japan: YHP, 1-59-1, Yoyogi, Shibuya-Ku, Tokyo, 151.

00208

HEWLETT hp PACKARD

THE AMATEUR SCIENTIST

Outdoor temperature, humidity, wind speed and wind direction are displayed on a panel

Conducted by C. L. Stong

If you have ever hankered to have the outdoor temperature, relative humidity, wind speed and wind direction displayed on a panel in your living room, it has been neatly worked out for you. David A. Sankey, an Air Force officer, has built a weather station that displays information of this kind in his home in Victorville, Calif. A similar station can be made by any amateur.

The sensing devices of Sankey's station are on a 20-foot mast about 300 feet from his house. The measurements are transmitted electrically to an instrument panel in the house by means of a cable. The cost of a station comparable to Sankey's will vary with the resourcefulness of the builder and the length of the cable, but it should not exceed $50. Sankey describes the construction and operation of his station as follows:

"The exterior sensing devices are mounted on a horizontal crossarm of wood that is bolted at its center to the top of a vertical pipe about 20 feet long. The pipe is supported by guy wires. The crossarm carries an anemometer to measure wind speed and a vane for indicating wind direction [*see illustration on opposite page*]. At a point about six feet above the ground the mast supports a louvered wooden box enclosing a pair of dual transistors and a cluster of 13 germanium diodes connected in series. The dual transistors function as a hygrometer of the wet-bulb, dry-bulb type. The 13 diodes and the associated circuit constitute a sensitive thermometer with an expanded scale for observing the size and movement of locally heated domains in the atmosphere.

"The anemometer is essentially a small windmill that generates an electric potential, which varies with the speed of the wind, and a voltmeter that can be calibrated for indicating velocity directly in knots or in miles per hour. Wind direction is sensed by a vane that is supported by the shaft of a self-synchronous motor. The shaft of a companion self-synchronous motor, which is mounted behind the instrument panel, rotates a pointer with respect to a compass card on the front of the panel.

"The operation of self-synchronous motors is analogous to that of meshed gears: when either motor is rotated by an external torque, its companion rotates in lockstep. In my apparatus the wind vane rotates one motor. The companion motor simultaneously turns the pointer at the panel to indicate the direction of the vane with respect to the cardinal points of the compass.

"The electric generator of the anemometer consists of a miniature direct-current motor of the type having permanent magnets that provide the required magnetic field. Such motors, which are available at hobby stores, cost about $1. Choose one that turns freely.

"Self-synchronous motors are widely advertised by surplus dealers at prices ranging upward from $5 per pair. The less expensive types are designed for operation at 26 volts and 400 cycles per second. At the reduced loading imposed by the wind-vane system the 400-cycle motors can be operated at 60 cycles per second without overheating if the rated voltage is lowered about 50 percent. The 26-volt motors of my system operate on 12-volt alternating current.

"The generator of the anemometer is mounted vertically inside a block of wood attached to the upper end of the crossarm [*see top illustration on page 124*]. Both the crossarm and the anemometer block are cut from standard two-by-four-inch pine stock. A hole that makes a snug fit with the motor is drilled approximately three-quarters of the way through the block near the long edge. A drill bit of smaller diameter is then used for extending the hole through the block. The smaller diameter is selected to make a press fit with a rod of nylon.

"A hole is drilled through the rod axially to make a sliding fit with a length of 3/16-inch brazing rod that serves as a shaft to support the impeller of the anemometer. A rectangular transverse opening is cut through the block so that it extends downward about an inch or so from the lower edge of the nylon bearing. The rectangular opening provides access to the shaft of the motor. Insert the motor into the larger hole and couple its shaft to the brazing rod by means of rubber tubing.

"The impeller assembly consists of three cups supported by equally spaced radial spokes fixed to a central hub. The cups were made by cutting off the spouts of aluminum funnels about 1/2 inch from the apex of the cone. The remaining portion of each spout was hammered to a point and sealed with epoxy cement. The hub is a disk of plastic covered with a jar lid for protection from the weather. A snugly fitting washer is fastened with epoxy to the shaft at the upper end of the nylon to act as a thrust bearing.

"The spokes that support the cones were made of brazing rod. The inner ends of the spokes were bent at right angles for insertion into holes of the hub, which are spaced at intervals of 120 degrees. The outer ends of the spokes were threaded for fastening to the cones with nuts. After the assembly has been mounted on the vertical shaft the cups are aligned by bending the spokes as required to make the impeller run true. The leads from the generator are brought out through a small hole in the side of the supporting block. Finally, the supporting block is screwed to one end of the crossarm.

"The wind vane consists of an aluminum tube 3/8 inch in diameter and about nine inches long. It supports a tail of balsa wood at one end and a counterweight at the other. The assembly is mounted to the vertical shaft of a self-synchronous motor by means of a collar that is both wired and cemented with epoxy to the aluminum tube at its balance point [*see bottom illustration on page 124*]. I bought the collar from a distributor of bearings and similar machine parts. The tail was cut from a 1/8-inch sheet of balsa wood. The wood was

protected with a coat of sealer and two coats of enamel. It was slipped into a saw kerf at one end of the aluminum tube and fastened with epoxy. A carriage bolt that makes a snug fit with the opposite end of the aluminum tube shifts the balance to within about three inches of the weighted end. To protect the upper bearing of the motor from dust and moisture I cemented the screw cap from a bottle to the bottom of the collar. The center of the cap was drilled to admit the shaft of the motor.

"The motor is clamped inside a conduit splice: a short length of tubing fitted with four setscrews. The device is available from dealers in electrical supplies. The bottom screw of the splice clamps the assembly to a cylindrical plug of wood that is fastened to the crossarm with a screw.

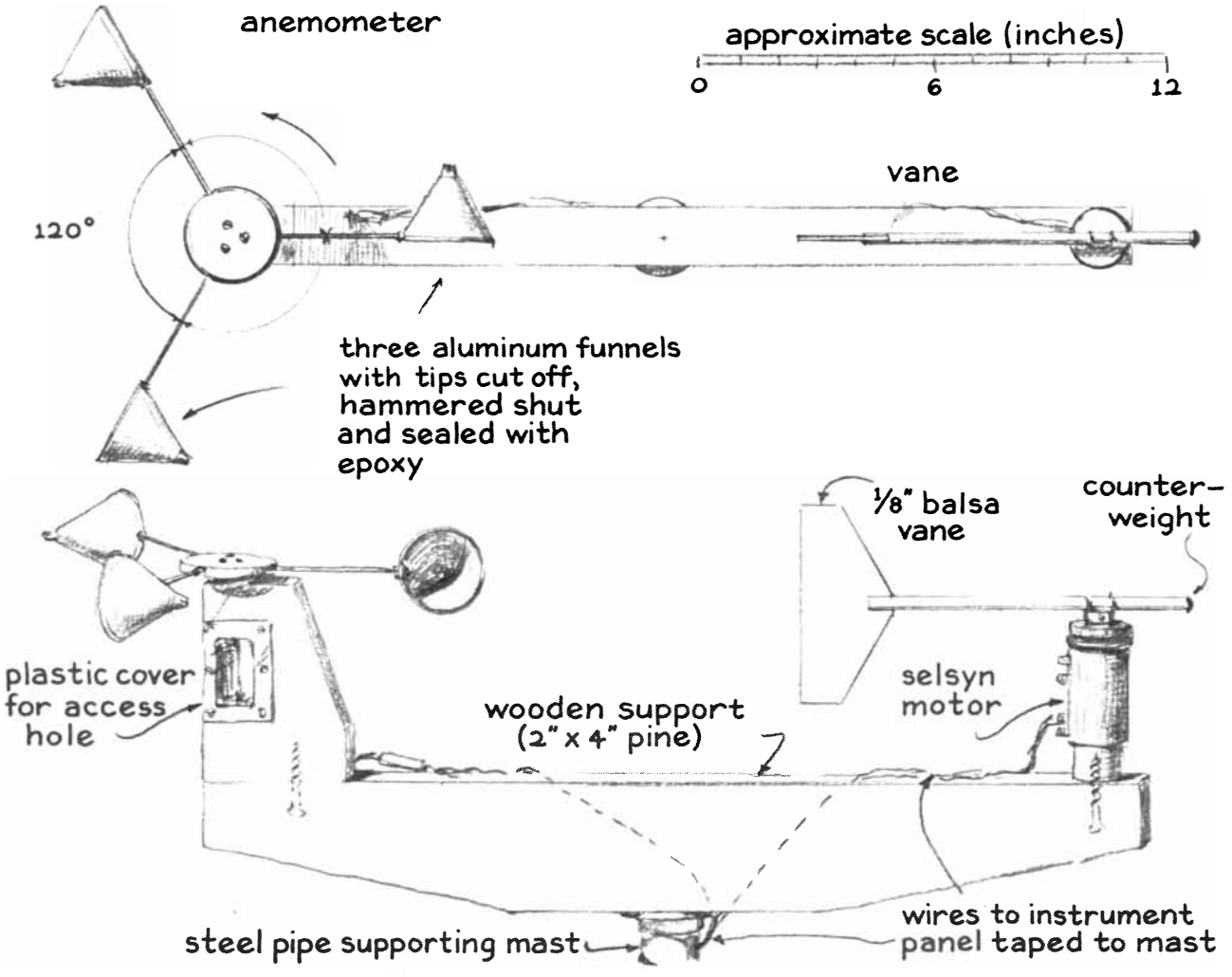
"The second setscrew from the bottom of the splice was removed. Leads from the motor were brought out through this hole. One side of the motor shaft was flattened with a file. The setscrew of the collar that supports the arm of the vane bears against the flattened side of the shaft. Hence the vane can be removed and replaced without disturbing the orientation of the vane with respect to the pointer on the instrument panel.

"The instrument panel can be improvised according to the tastes of the individual. The self-synchronous motor of the wind-vane system is mounted to the rear of the panel with its shaft extending through the front, where it supports the pointer [*see top illustration on page 127*]. To the panel must also be mounted at least one meter of the D'Arsonval type capable of measuring current through the range from 0 to 25 microamperes. The panel can include a rotary switch for connecting the meter to the anemometer circuit or to each of the three thermometers to be described. If cost is not crucial, each circuit can be provided with its own meter. The panel is completed by installing a transformer on the rear surface for reducing the power-line potential of 120 volts to 12 volts (alternating current). I also installed an on-off switch and a pilot light for convenience.

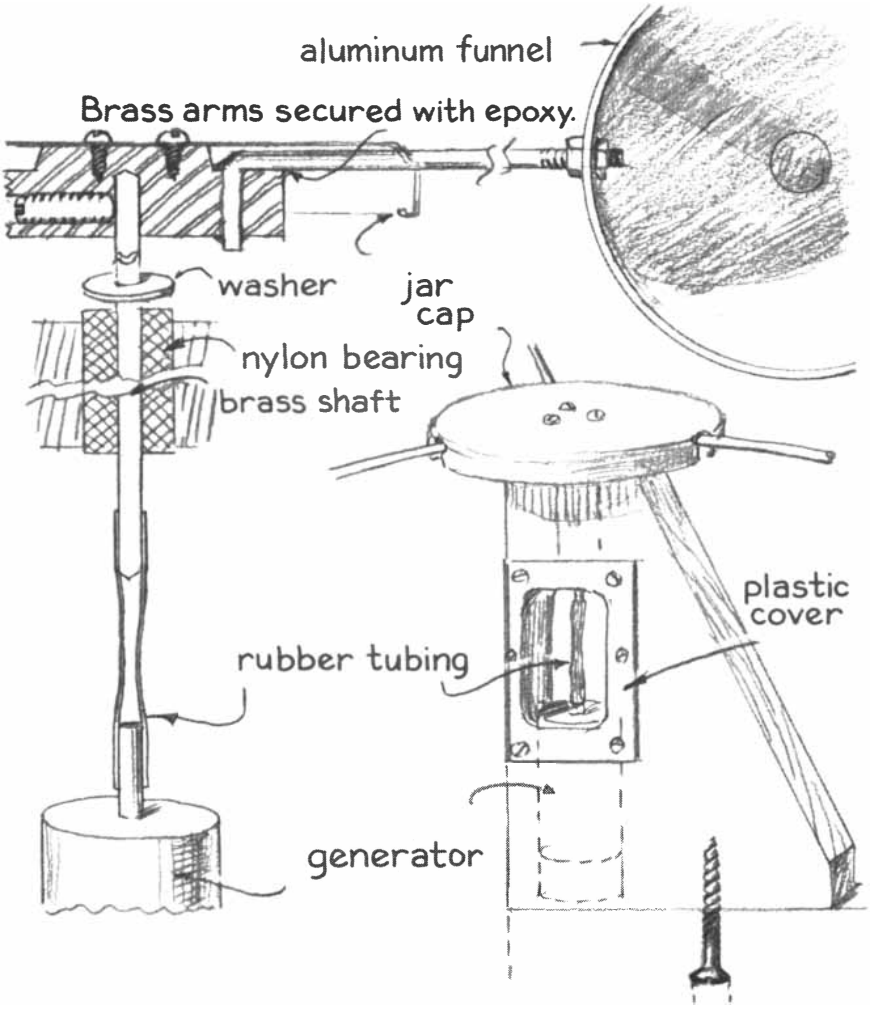
"The cable that connects the instrument panel to the anemometer and wind vane must contain seven insulated conductors—two for the anemometer and five for the wind vane. Each of the self-synchronous motors has five leads: one group of three and one pair. All leads are identified by either numerals or a color code. Connect all leads of one motor to the corresponding leads of the second motor by means of the cable. All cables are fitted with jacks and plugs for easy connecting and disconnecting.

"To energize the circuit connect the 12-volt output of the transformer to the pair of interconnected leads. When power is applied, each motor will rotate a fraction of a turn and stop. Thereafter when the shaft of either motor is rotated, the other motor will rotate in lockstep.

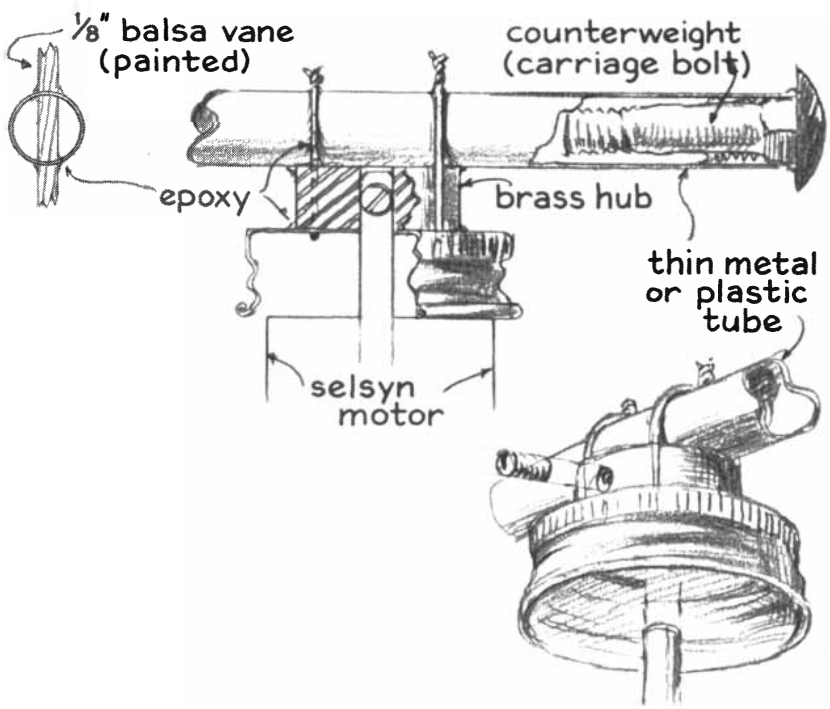
"To align the system, loosen the set-



David A. Sankey's anemometer and wind vane



Details of the anemometer



Details of the wind vane

screw that clamps the pointer to the shaft of the motor, turn the pointer to the position on the compass rose that corresponds to the direction in which the wind vane is pointed and tighten the setscrew. Some self-synchronous motors can lock into step at intervals of 90 degrees or 180 degrees. When motors of this kind are used, the alignment between the wind vane and the indicating pointer must be checked each time the power is turned off and on.

"The anemometer can be calibrated most reliably by making a second anemometer of the pendulum type. The pendulum anemometer serves as the calibration standard. The construction is quite simple. It involves suspending a ping-pong ball in the wind by a length of thread and measuring with a protractor the angle the thread makes with respect to the vertical when the ball is forced from its equilibrium position by the wind. Wind speed is calculated by means of a simple formula that takes the angle into account [see "The Amateur Scientist"; SCIENTIFIC AMERICAN, October, 1971].

"To make the calibration, remove the crossarm assembly from the mast and remount it temporarily on a support at eye level. On a breezy day connect the output of the anemometer to the meter with leads of convenient length. Insert a variable resistor of a few thousand ohms in one of the leads. Measure the wind speed with the pendulum anemometer. Increase or decrease the variable resistor to set the pointer of the meter at a desired point on the scale.

"For example, assume that the scale of the meter is graduated from 0 to 50 units and that the pendulum anemometer indicates a wind velocity of 20 miles per hour. The meter can be calibrated to indicate wind speed directly by setting the pointer at 20 units by means of the variable resistor. In general the potential developed by small generators having permanent magnets varies directly with the speed at which they turn. For this reason the anemometer can be calibrated with sufficient accuracy by checking a single point against the standard. A meter that is graduated in inconvenient units can be used with a graph made by plotting wind speed against the units of the scale.

"The optimum size of the variable resistor will vary with the characteristics of the generator and the sensitivity of the meter. Typically it will range from 1,000 to 20,000 ohms and must be determined experimentally. After the calibration has been made measure the ex-

SCIENCE/SCOPE

NASA's Earth Resources Technology Satellite, scheduled for launch this summer, will carry an experimental multispectral scanner developed by Hughes. The scanner's single optical system will record signal data in four separate bands of the electromagnetic spectrum and convert the light emissions into photo-like images. The resulting "signatures" of the solar energy emitted by agricultural crops, forests, and rivers will indicate their environmental health. The scanner will "see" a swath 100 nautical miles wide during each polar orbit over the U.S.

NADGE is the first major real-time air defense system to be implemented with a high-level compiler language (JOVIAL) using two Hughes-built processors in a multiprocessing mode. Hughes developed software programming for the 37 computer sites in the NADGE (for NATO Air Defense Ground Environment) system around common modules that can be adapted to solve problems peculiar to the various locations. Eighteen of the computer sites have been tested and accepted without any delay due to software programming problems. The entire NADGE system is scheduled for completion by the end of the year, except for some site work in Greece.

The first long-life hydrazine thruster systems, which Hughes developed for NASA's ATS-4 and 5 satellites, have proved superior to the conventional hydrogen peroxide thruster for making the radial and axial corrections that keep a synchronous satellite on its precise orbital station. The hydrazine thruster is safer, more reliable, longer lived, easier to restart in space, and less costly. Two new hydrazine engines, in 1- and 5-lb. thrust, are now available. Hughes is using the 1-lb. thruster on Canada's Anik 1 domestic communications satellite.

The U.S. Navy's AWG-9 weapon-control system which launches the Phoenix missile, both built by Hughes for the F-14 fighter, has a "look down" capability that enables it to pick out moving targets from the ground clutter that normally obscures conventional radar signals. A Phoenix missile demonstrated its ability to combat the anti-ship missile threat when it was launched from a test platform at 10,000 feet and hit a cruise-missile target flying at 800 feet, while another launched from 29,000 feet "killed" two targets flying close together at 10,000 feet. The AWG-9 can also launch the F-14's Sparrow and Sidewinder missiles and direct the firing of its M-61 Vulcan 20mm cannon.

Hughes needs software/programmer specialists with experience in software systems, real time applications, and technical data management systems. Also: business programmer/analysts with IBM 370/165 OS/MVT/ASP experience. U.S. citizenship required. Please write: Mr. R. S. Roth, Hughes Aircraft Company, P.O. Box 3310, Fullerton, CA 92634. An equal opportunity M/F employer.

New products from Hughes include a low-cost one-watt CW argon laser featuring the same sealed, cylindrical head as Hughes' "Hip Pocket" helium-neon lasers; the power supply is air cooled, permitting a compact 7x16x19-inch design for standard rack mounting . . . and a miniaturized scan converter designed for video data processing in a wide range of information display applications; it offers high speed and high resolution, as well as selective erasure and rewriting capability.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

New Horizons in Astronomy

JOHN C. BRANDT, NASA-Goddard Space Flight Center and University of Maryland, and
STEPHEN P. MARAN,
NASA-Goddard Space Flight Center

1972, 496 pages, approximately 320 illustrations (11 in full color), $12.50

A book that discusses the relevance of astronomy to the basic earth, life, and physical sciences—and indeed to the very origin and nature of life on Earth. While emphasizing the importance to every citizen of a broad understanding of science, the authors also treat specific developments such as pulsars, quasars, and the Apollo Moon studies at unusual length for an elementary survey.

GALACTIC ASTRONOMY

DIMITRI MIHALAS, Princeton University Observatory.
With the Collaboration of PAUL McRAE ROUTLY,
American Astronomical Society

"Welcome to this new addition to our libraries. Not only is it excellent, but to my knowledge it is unique—a general intermediate-level book on the astronomy of our galaxy. We may wonder why such a work has not appeared in the 40 years or so since galactic research began its rapid maturation. . ."
—Stanley P. Wyatt in *Sky and Telescope,* May 1969

1968, 257 pages, 113 illustrations, 42 tables, $10.00

STELLAR ATMOSPHERES

DIMITRI MIHALAS, Yerkes Observatory, The University of Chicago

"Required reading for those who seek an understanding of the physics of radiative processes in stellar atmospheres . . . Dimitri Mihalas' skillful, crisp, concise style permits him to embrace a vast range of involved mathematical theory in a reasonable amount of space. . . . For many years to come Mihalas' *Stellar Atmospheres* will occupy an honored place in astrophysical literature."
—*Sky and Telescope,* September 1971

1970, 463 pages, 82 illustrations, $16.00

RADIO ASTROPHYSICS

Nonthermal Processes in Galactic and Extragalactic Sources

A. G. PACHOLCZYK,
Steward Observatory, The University of Arizona

"It is rare in such a rapidly developing field as radio astronomy to find a new book that does not promise to go quickly out of date, but A. G. Pacholczyk of Steward Observatory appears to have produced a text that will be useful for many years."
—*Sky and Telescope,* December 1970

1970, 269 pages, 64 illustrations, $13.00

INTRODUCTION TO THE SOLAR WIND

JOHN C. BRANDT,
NASA-Goddard Space Flight Center, Greenbelt, Maryland

"Intended as an intermediate-level reference text, the book provides a comprehensive but highly readable account of the evolution of present-day ideas about the interplanetary medium and its relation to solar activity. . . . I found one of the most rewarding parts of the book to be the opening chapter, in which Brandt traces the history of early thoughts and theories of the interplanetary medium. In my opinion, even most experts in the field of space physics would benefit from reading this brief account of little-known facts."
—Roger A. Kopp in *Science,* September 18, 1970

1970, 199 pages, 85 illustrations, $10.00

FRONTIERS IN ASTRONOMY

Readings from SCIENTIFIC AMERICAN

With Introductions by
OWEN GINGERICH, Harvard University and
Smithsonian Astrophysical Observatory

"In spanning the period from 1956 to 1970 this collection provides important perspective in a rapidly moving field. The selection has been well made; the authors represented are well known. . . . The book is particularly suited to browsing, and the excellent illustrations are themselves worth its price. It can be warmly recommended to anyone with an interest in astronomy and astrophysics."
—*American Scientist,* March-April 1971

1970, 370 pages, 339 illustrations (170 with color), cloth $11.00, paper $5.75

From your bookseller or Dept. J
W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

---

act resistance with a volt-ohm meter and replace the variable resistor with a fixed resistor of the same value. The crossarm assembly can then be permanently mounted on the mast.

"Incidentally, the site of the mast is important. The anemometer and wind vane should be in an open area or on a roof, well removed from buildings, trees and similar obstructions. The instruments should be at least 10 feet above the highest neighboring obstruction.

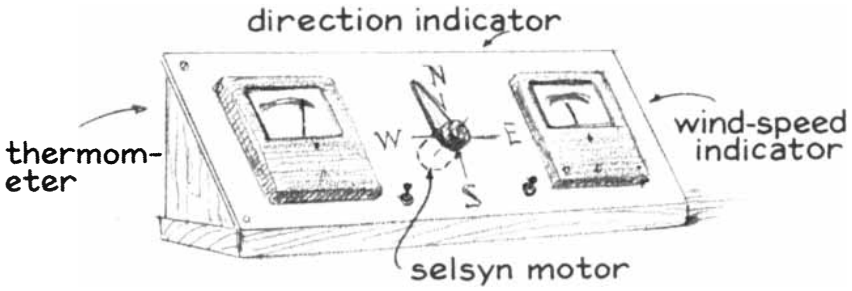
"The electronic thermometers used for measuring relative humidity and for detecting the movement of local air masses by observing differences in temperature are of two types. The temperature probe of the hygrometer is a pair of dual transistors that can be located several hundred feet from the associated circuitry. The output voltage of each dual-transistor circuit varies in direct proportion to the temperature. The circuit [*see bottom illustration at right*] was developed by T. C. Verster of the National Research Institute for Mathematical Sciences in the Republic of South Africa. Current in the collector circuit of one transistor of the dual pair is made to exceed the current in the collector circuit of the companion transistor by a ratio of more than 10 to one. If the ratio is fixed, the difference between the base-to-emitter voltages of the two transistors is proportional only to the absolute temperature of the transistors. The operational amplifier maintains the current ratio by means of negative feedback.

"The device is remarkably simple to calibrate. To make the calibration connect a voltmeter between the wiper arm of the 1,000-ohm potentiometer (R_a) and the chassis. Apply power to the circuit. The reference voltage developed by the Zener diode (1N3157) appears across the voltmeter. Adjust the reference potential to −273 millivolts by adjusting the 1,000-ohm potentiometer. (A millivolt is one thousandth of a volt.) Insert the transistor and a mercury thermometer in an oil bath at room temperature. Adjust the 750-ohm potentiometer (R_b) to the point where the meter indication in millivolts matches the temperature in degrees Celsius as indicated by the thermometer. Verster recommends the use of a digital voltmeter with an ungrounded input. A conventional meter of the D'Arsonval type will work, although it is less convenient than the digital voltmeter to read. The calibration need not be checked at more than one reference point.

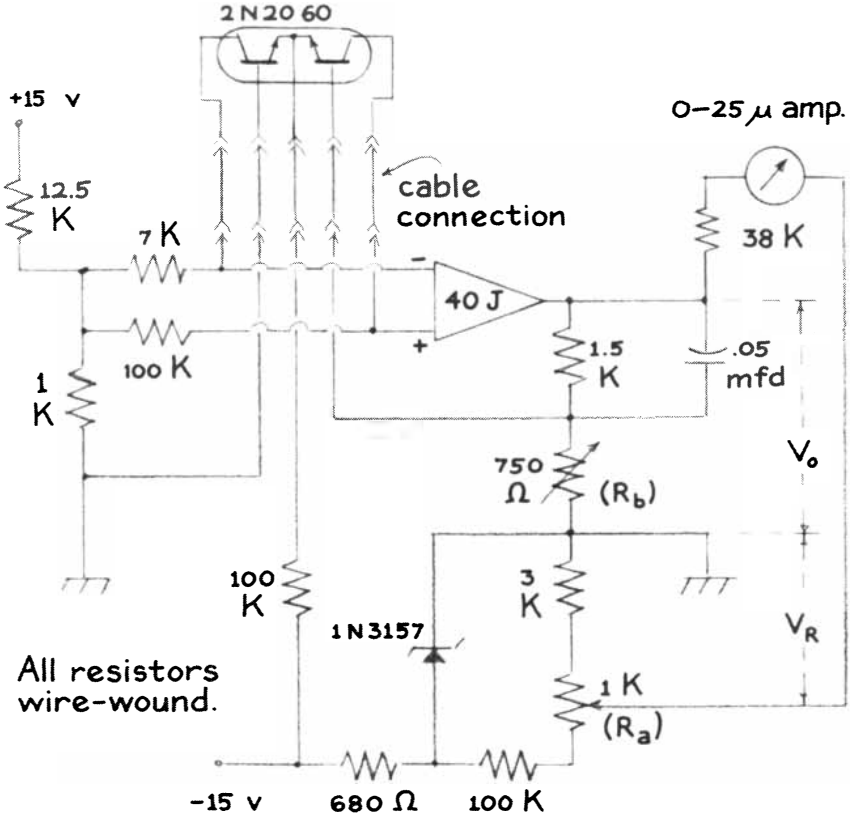
"Both transistors are cleaned, coated lightly with epoxy and attached to shielded, five-conductor cables. They are suspended vertically side by side about an inch apart and half an inch from a supporting panel of wood. The panel can consist of a piece of quarter-inch plywood five inches wide and 10 inches long with a spacer strip half an inch thick, two inches wide and five inches long glued across the upper end. The cables are supported by the spacer strip. A tubular wick of the kind used in hygrometers is slipped over one of the transistors and tied in place with thread. The wicks are available from dealers in scientific supplies for $2 per dozen. Wicks should be changed about four times a year.

"A glass cistern is clamped to the panel with its opening immediately under the transistor. Glass cisterns are also stocked by dealers in scientific supplies and cost about $2 each. Fill the cistern with distilled water and insert the wick. The companion transistor functions as the dry bulb of the hygrometer.

"The temperature probes can be housed in a box that has louvers large enough to ensure the free circulation of air. The box should be painted white. It can be placed at any outdoor site, preferably one that is shaded from the sun. The circuits can be energized by a pair of 12.6-volt batteries such as the Mallory Type TR-289 mercury cells or, preferably, by a regulated 15-volt dual power supply of the type previously



Indoor instrument panel of the weather station



Circuitry of the transistor thermometer

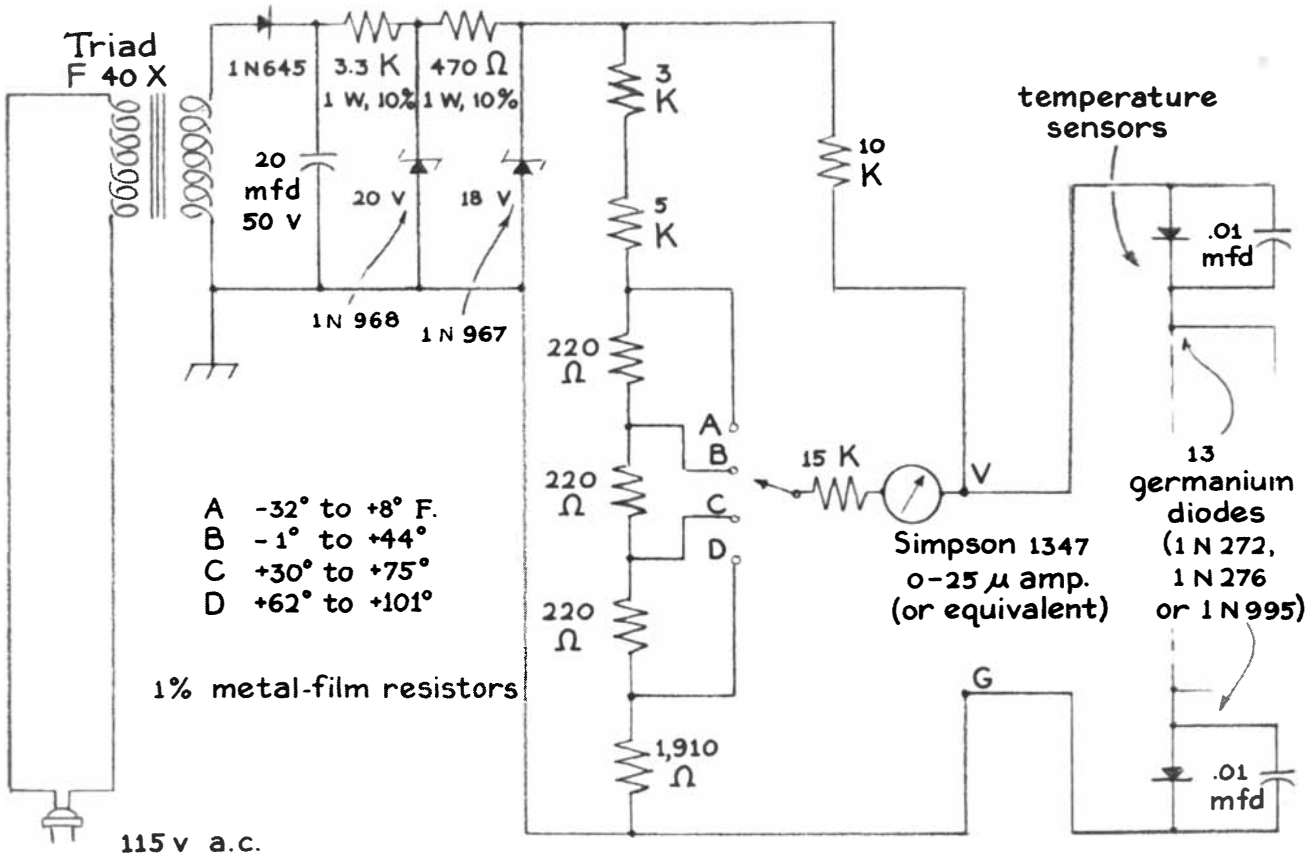
described in these columns [see "The Amateur Scientist"; SCIENTIFIC AMERICAN, May, 1970]. Relative humidity is determined by measuring the temperature as indicated by the wet and dry probes and referring to a table published by the U.S. Weather Bureau in bulletin No. 1071. The table is reproduced in standard references such as *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Company). Incidentally, the accuracy of the hygrometer can be improved by installing a small electric fan near the wet bulb to encourage evaporation of water from the wick. Let the fan run for about two minutes before making a reading.

"The sensing probe of the expanded-scale thermometer used for observing micrometeorological effects consists of a string of 13 germanium diodes connected in series. The diodes can be assembled in the form of a compact cluster about 1½ inches square by bending the leads of each unit to form a *U* and pushing them through a rectangular pattern of holes drilled in a sheet of thin Formica or a similar plastic. The electrical resistance of the diodes varies in inverse proportion to the temperature. The potential drop across the string varies by a total of one volt from −24 to 38 degrees C. (−11 to 100 degrees Fahrenheit).

"The diode assembly constitutes one of the four arms of a Wheatstone bridge [*see illustration below*]. The meter, which is connected across the bridge, indicates average temperature. The circuit, as developed by Douglas A. Kohl of Minneapolis, includes a rotary switch for observing temperature in four slightly overlapping ranges of approximately 40 degrees each. In effect the insertion of the switch magnifies the three-inch scale of the meter to 12 inches, thus increasing the accuracy with which readings can be made. Rapid fluctuations of temperature appear as a voltage of varying amplitude across the diodes. It can be monitored directly by connecting the input of an operational amplifier, such as the type designated 741C, to the circuit at points *V* and *G*. If desired, the output of the amplifier can be used for driving a chart recorder.

"The thermometer circuit is energized by a power supply that includes a step-down transformer for reducing the line potential to 26 volts. A diode converts the output of the transformer to direct current. The remaining portion of the circuit includes a pair of Zener diodes in tandem that deliver a constant potential of 18 volts to the Wheatstone bridge.

"The meter indications vary in direct proportion to the temperature. For this reason a calibration graph that relates temperature to meter indications can be drawn by measuring the temperature of the diodes at two points in each of the four ranges of the instrument and tabulating the corresponding meter indications. For example, set the rotary switch to the 62-to-101-degree range. With a mercury thermometer measure and record the temperature of the room and the meter indication. Transfer both the thermometer and the diodes to a warmer environment, such as the interior of a cardboard box that is heated by a 100-watt incandescent lamp. After the temperature stabilizes make the second set of readings and draw the graph. Lower ranges can be similarly calibrated by using the cooling and freezing compartments of a refrigerator. Some of the components of the thermometers, such as the dual transistor and the germanium transistors, are not stocked by distributors that cater to amateurs. They can be obtained from Semiconductor Specialists Inc. (P.O. Box 66125, O'Hare International Airport, Chicago, Ill. 60666)."

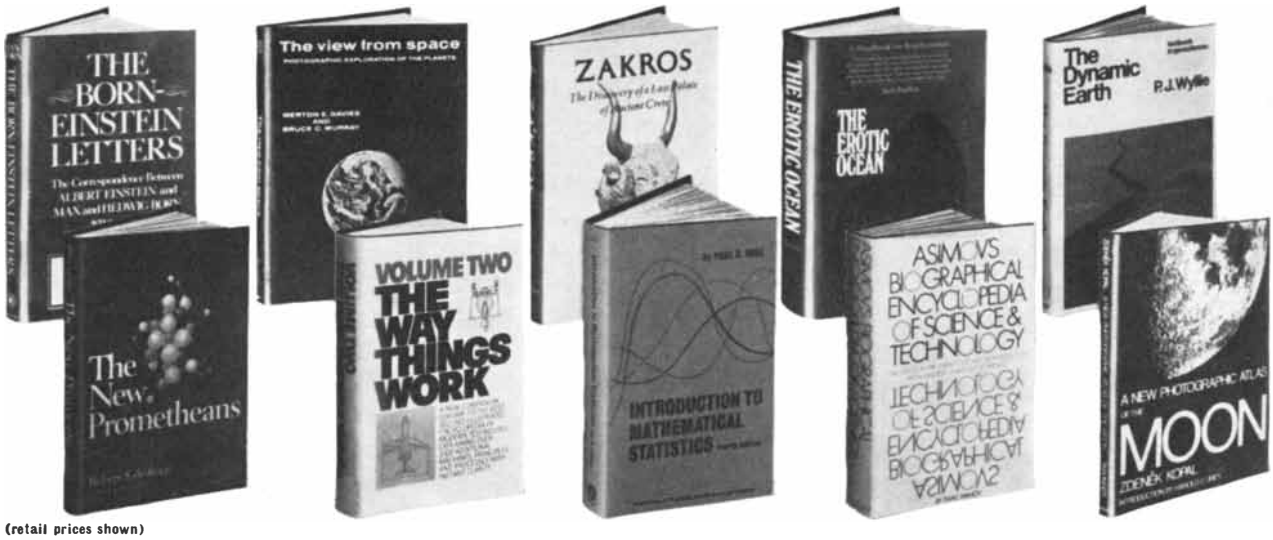


Wiring of the diode thermometer

You'll love roads you used to hate.

Mercedes-Benz
Mercedes-Benz of North America, Inc.

This summer, catch up on the ones that (almost) got away.



(retail prices shown)

88690. ZAKROS: The Discovery of a Lost Palace of Ancient Crete. *Nicholas Platon.* Illustrated step-by-step description of the excavation of a great palace of ancient Crete—as good as a well-written detective story. Illustrated. *Counts as 2 of your 3 books.* **$19.95**

36770. THE BORN-EINSTEIN LETTERS. The correspondence between Albert Einstein and Max and Hedwig Born, 1916-1955. With commentaries by Max Born. Translated by Irene Born. **$8.50**

73180. QUANTUM THEORY AND BEYOND. *Edited by Ted Bastin.* 21 of the finest minds in the world consider the restructuring of quantum theory — and how it is to be done. **$16.00**

34610. ASIMOV'S BIOGRAPHICAL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY. New Revised Edition. *Isaac Asimov.* **$12.95**

56580. INTRODUCTION TO SPACE SCIENCE. *Robert C. Haymes.* Nuclear clocks, quasars, the origin of the solar system, and the Van Allen belt are among hundreds of topics covered in this brilliant and lucid overview. Follows in the tradition of Ball, Jeans, Eddington, Shapley, and Hoyle. **$14.95**

56610. INTRODUCTION TO SCIENTIFIC COMPUTING. *Gabriel A. Pall.* Present day computers are not just overgrown adding machines; rather, they are full partners of scientists and engineers. Here, an IBM expert provides the know-how needed to use computers to maximum advantage. *Counts as 2 of your 3 books.* **$19.95**

70510. PROBLEM-SOLVING METHODS IN ARTIFICIAL INTELLIGENCE. *Nils J. Nilsson.* Basic approach to one of the most stimulating new computer fields: having machines perform tasks normally requiring human intelligence. With many examples. **$10.50**

69270. A PICTORIAL HISTORY OF AMERICAN MINING. *Howard N. and Lucille L. Sloane.* The swindles and strikes, the colorful personalities and tragic disasters — are re-created in a meticulously researched, illustrated volume. **$12.50**

64580. THE NEW PROMETHEANS. *Robert S. de Ropp.* Perceives a continuing battle between the creative and destructive forces of science and applies this thesis to the great scientific developments of the recent past. **$7.95**

The Library of Science

invites you to select any 3 books

(values to $48.50)

all for only $3.95

if you will join now and accept only 3 more books in the next 12 months at membership discount prices

79940. THE SPECTRA AND STRUCTURES OF SIMPLE FREE RADICALS. *Gerhard Herzberg.* Spectroscopy is the most powerful tool for studying these chemical will-o'-the-wisps and this new volume by the master is an important addition to the literature. **$11.00**

86510. THE WAY THINGS WORK. Vol. Two. Ski bindings, submarines, artificial kidneys—you name it—you'll find the inside, expert story about what makes everything tick. **$9.95**

85830. THE VIEW FROM SPACE. "Chock full of solid substance for the serious student of space photography yet has an historical vein and a pleasant conversational style. Timely, authoritative..."—James A. Van Allen. **$14.95**

42970. THE DYNAMIC EARTH. *Peter Wyllie.* The 1960s saw the rise of global dynamics, a revolutionary and all-embracing theory of geology. Here is an equally revolutionary book to explain the new global tectonics. **$17.50**

45600. THE EROTIC OCEAN. *Jack Rudloe.* Everything marine enthusiasts need to locate, preserve, and identify oceanic specimens. This illustrated volume is an important contribution to conservation as well as a guide to scientific beachcombing. **$15.00**

63860. THE NATURE AND GROWTH OF MODERN MATHEMATICS. *Edna E. Kramer.* Ingenious new approach to the history of developing mathematical concepts. Lucid, brilliantly organized, even dramatic, this is a real find. *Counts as 2 of your 3 books.* **$24.95**

56260. INTRODUCTION TO MATHEMATICAL STATISTICS: Fourth Edition. *Paul G. Hoel.* Extensive rewriting of a great classic. Mathematically rigorous yet filled with concrete problems and practical applications. **$11.50**

56540. INTRODUCTION TO MATRIX ANALYSIS, Second Edition. *Richard Bellman.* Matrix theory is central both to computer operation and the understanding of its associated mathematics. The most authoritative work on the subject. **$14.75**

64910. NUMBER WORDS AND NUMBER SYMBOLS. *Karl Menninger.* An oversized, handsomely illustrated, 480-page volume filled with unusual lore and amazing insights into number words and the universal concepts they represent. **$14.95**

36290. BIOPHILOSOPHY. *Bernhard Rensch.* Recent revolutionary discoveries in biology form the basis for an unprecedented natural philosophy of life constructed by the brilliant German biologist-philosopher. **$12.50**

56350. INTRODUCTION TO GEOMETRY, Revised Edition. *H. S. M. Coxeter.* Urbane, unconventional, witty — a modern classic. **$10.95**

64500. A NEW PHOTOGRAPHIC ATLAS OF THE MOON. *Zdenek Kopal.* Breathtakingly illustrated with more than 200 unique photographic plates, here is a lunar Baedeker that will allow you to explore the Moon first-hand. *Counts as 2 of your 3 books.* **$20.00**

MEMBERSHIP APPLICATION

THE LIBRARY OF SCIENCE, Riverside, N.J. 08075

2-646

Upon acceptance of this order, please enroll me as a member and send the three books I have indicated. Bill me only $3.95 for all three, plus postage and handling. If not delighted, I will return all books within ten days and this membership will be cancelled.

As a member, I need accept only three more selections during the next 12 months at reduced member prices, plus postage and handling. Savings range up to 30% and occasionally even more.

I understand that I will receive free advance Reviews which fully describe each month's Main Selection and Alternates. If I wish the Main Selection, I need do nothing; it will come to me automatically. If I do not wish to receive the Main Selection, or I wish an Alternate Selection, I will advise you by using the convenient reply card provided with the monthly Reviews. I understand that I may choose a bonus book for every four selections purchased. (Introductory offer counts as first selection toward a bonus book.)

Send no money. Members are billed when the books have been received.

(Offer good in Continental U.S. and Canada only.) Prices slightly higher in Canada.

3 Books for $3.95. (write in numbers): Some expensive books count as 2 selections.

_____ _____ _____

Name _____

Address _____

City _____ State _____ Zip _____

BOOKS

War from the air, Newton's Principia, the wrapping of a cliff and other topics

by Philip Morrison

THE AIR WAR IN INDOCHINA, edited by Raphael Littauer and Norman Uphoff. Air War Study Group, Cornell University. Beacon Press ($3.95). In Theodore Roosevelt's day Yankee know-how, compounded of bold and imaginative machine design, epidemiological understanding and adept military management, moved in a decade several hundred million cubic yards of tropical rock and dirt to dig the Panama Canal. A sharp track remains across the green isthmus. It is a long-lived engineering and economic success, however faulty our title to the land. Today, in another tropical time zone, the same American know-how, far more advanced technically and of even more doubtful legality, has heaved up several Panama Canals' worth of soil and subsoil. Scattered across the green land of Indochina are millions of bomb craters, craters crowded on craters as they are in the lunar highlands. No canal results from this investment, only an enduring pattern of infertile circular algal ponds, "difficult to drain," "virtually unsuitable for cultivation," mosquito havens that General Goethals would not have tolerated.

The system that conducted most of this large-scale antiengineering work is a small part of the United States Air Force. Indeed, the task has absorbed about half the lift over a few years of 50 to 100 B-52's, eight-jet intercontinental bombers designed to carry thermonuclear weapons at Mach .9. They are now operating mainly from coastal Thailand bases with "iron bombs," about 100 per load, each bomb holding on the average 500 pounds of high explosive and an impact fuze in a light shell. Almost every drop yields a crater, typically 30 feet across and 15 feet deep. Once in a while a real landslide follows a highland salvo, or a few rice farmers or a three-quarter-ton truckload of enemy ammunition intercepts some fragments of the shrapnel.

A couple of thousand chemical workers back in Tennessee make the explosive, which is poured into the cheap stamped casings, shipped to the Thailand coast, fork-lifted into the big bomb bays and hauled over the landscape to be dropped by loran coordinate on nowhere. It is "counterinsurgency from 30,000 feet." Each of the eight million or so B-52 craters claims about $400 from the taxpayer, making no allowance for amortization, training, research and development. Smaller aircraft have made a few million additional holes at higher cost. At most one or two of the B-52's have been lost; they fly high enough to avoid antiaircraft gunfire, and they try to stay above the range of the Russian missiles. Encapsulated within a cruel war, this extraordinary visitation of high technology on a developing country deserves the ironic attention of poets; it is too intense for the prosy historian.

Published photographs of the cratered countryside are not common, although some were published in the last issue of this magazine. There are none in this graph- and table-filled book, the remarkable output, at once engaged and cool, of a group of 20 students and faculty members, mostly from Cornell University. They prepared their research plan in the spring of 1971, worked hard and expertly and produced this study with data as late as February, 1972. It is the first book-length analysis of the air war on Indochina, a kind of warfare whose remoteness and institutionalization have provided for the executive branch of our government an agent almost beyond restraint.

The authors have collected the history, the facts, the law, the detailed circumstances and the costs in dollars, lives and land of the air war. Their sources are published and unpublished; the unpublished data come from direct interviews with more than 80 consultants from the Department of Defense, Congress, the press and the universities. They have considered the issues widely; ecology, the constitution, the Algerian parallel and international law are all treated. The consistency, candor and modesty of the presentation give the book lasting value. The authors do not conceal their dislike of the war, but above all they seek to understand it, the better to limit its power.

Decision in Indochina still lies, as it always has, in the political impact of widespread infantry combat, but the Cornell group sets forth in one vivid page of graphs the political intention of the air war. The curves show the rise and decline of bomb tonnage. They present not only the total for all Indochina but also the totals for each of five distinct campaigns of the air war: over South Vietnam, North Vietnam, northern Laos, Cambodia and the "Ho Chi Minh trail." The overall total of air-dropped explosive—a tonnage three times that of World War II and six times that of the Korean war—peaked during the election year of 1968. Laos peaked later, and it still remains high. The campaign on the Ho Chi Minh trail grows steadily; bombing there was higher in 1971 than in any other year. Vietnamization plainly included the reduction of draft calls at home, a maximum investment in air interdiction and a renewed effort to reduce the flow of supplies to the other side—an effort that appears to be no more successful than it has been for a decade. Air-power theorists like to reckon what part of the munitions flow is destroyed; their critics observe that what counts is how much gets through. Heavier attacks destroy more each year because each year more is being shipped.

The most ominous narrative in the book describes the innovative attacks conducted by fast fighter-bombers and night-flying gunships: big, slow aircraft with image tubes to sight guns able to fire up to 6,000 40-millimeter shells a minute. Along the Ho Chi Minh trail are seeded thousands of sensors, each an integrated electronic module, sniffing for sounds, odors, infrared glow, ground vibrations, even the scent of urine. These battery-powered devices pulse their automatic signals to relay points in the line of sight, sometimes on mountaintops,

sometimes in high-flying planes, sometimes even to light unmanned aircraft droning automatically overhead all night.

Back in Thailand at a headquarters called Commando Bolt the computer records, collates and finally displays a line of light—"the worm"—that crawls across the screen, mapping some distant convoy of trucks moving forward by night. "From there the battle becomes academic." Coordinates follow and the most suitable aircraft and weapons are chosen as the clever software optimizes and compromises routes and targets. The assessment officer directs the operation by radio from his air-conditioned trailer hundreds of miles from the battle. He "determines tactical necessity." Maybe the worm is cut and stops crawling. "War and war games become much the same."

Or so the proposal reads. The Air Force and its symbiotic industries are imaginative, ingenious, well organized and lavish. Their model of the battlefield, so capital-intensive, so rational, so risk-free, depends heavily on what euphemists call a "permissive environment," that is, no matching opposition. That armed peasants cannot compete is axiomatic too. They are now, however, enforcing a less permissive style. Radar-controlled antiaircraft guns appear along the trail, and some gunships go down. Surface-to-air missiles (SAM's) are set up. Mimeographed sheets circulate from experienced infantry officers somewhere in General Giap's entourage, and some trucks act like stage armies, loading an attentive sensor with delectable targets that may exist only on our magnetic disks. No doubt about it: sophistication is also a blueprint for escalation. We found the guerrillas fighting with crossbows and purchased M-16 rifles; we observe them now with SAM's, howitzers, tanks and six-inch rockets.

These elements are not the point of the story, although they are important. Two conclusions stand out. First, the steady advances in weapons technology increasingly free their designers and operators from moral or humane concerns, producing "sophisticated barbarism on an ever increasing scale." "An automated war might soon fade from the public consciousness and become *institutionalized.*" Legal and moral restraints are attenuated. The combat personnel are self-selected, pretrained, highly professional, supported out of a traditional defense budget. One day bomb delivery will be automatic too, and there will be few pilot losses. Industrial support itself is cut off from public view. Dow gave up napalm, but some small specialty plant makes it still in a community where the hundreds of jobs seem important and the gouts of flame are distant.

Second, the critic is bound to remark that the Air Force, one of the largest and most energetic organizations in the world, is threatened today by a kind of bankruptcy of purpose. Strategic bombardment by air is all but obsolete; the 1,000 targetable Air Force missiles on land are obsolescent, perhaps even provocative and destabilizing. Air superiority over the battlefield cannot justify the traditional Air Force scale: full parity with the Army and the Navy. An Air Force victory in Laos and Vietnam would restore a military future in other peasant countries. Hence the attractiveness of the new, expensive Orwellian war by sensor, computer, drone and mass-produced firepower. A humane future for technology faces no greater peril than the growth of the "electronic battlefield." It is uniquely the product of the American technical community, and it is surely the duty of that community to perceive and then to explain how profoundly irrational and how suicidal in the large this terrifying divorce between means and ends is. The essential feedback loops—individual and moral, social and political—that over the centuries have enforced in most countries a crude proportionateness of violence are becoming lost in a new noise, in the pulsed flood of bits that arise from a false and narrow rationality. Those controls must be restored, indeed strengthened, in our thermonuclear times. Read the Cornell book as the air strikes expand. It may be a saving first step.

INTRODUCTION TO NEWTON'S 'PRINCIPIA,' by I. Bernard Cohen. Harvard University Press ($30). PHILOSOPHIAE NATURALIS PRINCIPIA MATHEMATICA: VOLUME I AND VOLUME II, assembled and edited by Alexandre Koyré and I. Bernard Cohen, with the assistance of Anne Whitman. Harvard University Press ($60). Johannes Kepler set down his three laws for the motions of the planets around their somehow magnetic sun in the decade after Galileo first explored distant worlds by sensory means with which men were not born. Some 50 years later, in 1673, in a Europe where more than one royal capital supported a well-established observatory and a scientific journal, Christiaan Huygens, in a treatise of real mathematical depth, showed how the centrifugal force in uniform circular motion must vary with speed and radius. Edmund Halley wrote that he "came one Wednesday to town" in January, 1683, to meet and discourse with "Sr Christ. Wrenn and Mr Hook" about orbits. Halley and his friends Christopher Wren and Robert Hooke could all see by simple algebra that an inverse-square attraction from a center would balance the Huygens force to yield Kepler's period-radius results, by now seen in the satellites of Jupiter and Saturn as well. According to Halley, Hooke claimed even more: "Upon that principle all the laws of celestiall motions were to be demonstrated, and that he himself had done it." Hooke and Halley never did succeed in the demonstration; indeed, possibly only Hooke, and not Halley or even Huygens, had a clear view of a central solar force rather than some simulating mechanical arrangement. Wren offered Hooke ("besides the honour") "a present of a book of 40 s.," but no demonstration came.

That August, it appears, Halley went to Cambridge to consult with Isaac Newton, who was Lucasian Professor there. Newton was 41 years old, known for his publications on light and color and "admired by the cognoscenti" for his depth in mathematics and dynamics, judged from lectures and manuscript. Newton had already got all the Londoners' results and had gone beyond their circles to real ellipses to extract all the Kepler laws five years before. He could not find his papers, and so he rederived the results and sent Halley the gist of the matter in a short paper (which we do not possess). Back came the excited Halley to visit Newton, and on December 10 Halley told the Royal Society he had seen "a curious treatise, *De Motu.*" By February a paper was available in London. (There are several confusing versions, one of which may be the same as Halley's.) The stimulus from London was strong and the rivalry with Hooke was real; within 18 months the Royal Society received a book manuscript of the *Principia.* (The Royal Society still has it.) It was 460 sheets, a clean final copy, with only minor revisions, written out by an amanuensis, Humphrey Newton (no relation), a physician and male midwife. The most celebrated text of modern science was in the printer's hands.

The volumes reviewed here are loving and learned labor over nearly two decades by two distinguished historians of science. (Professor Koyré of Paris died in 1964.) They extend to this classic of our science the depth of critical attention granted to the great poets and playwrights of every tongue. The center of

the work is a careful photographic reproduction of the last edition of the *Principia* Newton was to see, printed in his 84th year. To this Latin text is added at the bottom of each page every difference found in seven successive versions of the Latin work: the original printer's manuscript of 1686–1687, the two earlier editions, as printed in 1687 and in 1713, and the only four known copies (two for each edition) annotated with corrections in Newton's own hand.

The text changes amount to considerably more than a tenth of the whole over three editions; Newton added the results of new experiments, observations on falling bodies, on lunar positions, on refraction, the orbits of comets, on precession and on much more. He responds to criticism, mainly from his editors, refining and modifying definitions, proof and examples. He struggles a long time to clarify what was meant by the "quantity of matter." The theological and philosophical sections are perhaps the most reworked, because they were what held the most general interest, as they plainly still do for historians. "*Hypothesis non fingo*" appears in the fifth manuscript draft of one famous passage tacked on to the end of Book II in the second edition! Like other widely quoted material, this was a response to the philosophical critique by Leibniz.

Such a painstaking study of a single work is not for hurried readers, although everyone should at least look through the volumes. Happy the historian, steeped in the men and the times he studies, alert to every nuance, privy to dusty cartons of enigmatic mathematical scratchings, who can recall "how exciting it was to find on page 2 of Newton's interleaved copy [of the second printed edition] a slip of paper containing a longhand statement contrasting the Keplerian and Newtonian concepts of inertia." One can see a facsimile of that Latin slip in the *Introduction,* a humane and perceptive biography of the *Principia* and a delight for the scientific booklover. Another find is a list Newton prepared suggesting who should get 70 free presentation copies of the second edition. "The Czar 6 for himself & ye principal Libraries in Muscovy [Newton may have met the Tsar—it was Peter the Great—in London] ... young Monsr Cassini, De La Hire, Maraldi, & Varignon ... Mr Bernoulli at Bazil, Mr Leibnitz. ... In Millain the publick library."

The books are beautifully made and well illustrated. There remains a volume of commentary to come; it will present and analyze the key texts of Book I and of Book III with an English translation, a complete English table of contents, essays on the data used by Newton and the like. We can hope one day for still two more volumes, carrying on in the same style for the whole of the work. "I am not so rash as to promise to produce these myself," writes the Harvard author, whose warmth pervades the text even during its most meticulous sorties into detective scholarship.

The entire Newton text is going word by word onto magnetic tape, which other scholars can draw on and which will eventually yield a published analytical concordance of the *Principia.* By the tercentenary year of 1987 the scholars will have little left to do, although it is doubtful that we shall have fully charted that lonely voyaging mind.

CHRISTO, text by David Bourdon. Harry N. Abrams, Inc. ($25). Christo Javacheff is a New York artist of worldwide reputation. He was born in Bulgaria, studied art in Sofia and stage design in Prague and went to Paris in 1958. Among the many artists whose burning focus is novelty of concept and form, who regard the act of execution as being fully as important as the completed work, who use ephemeral materials undreamed of by the old masters, Christo—as he is called in the old complimentary tradition—makes about the freshest, most revealing, scrupulously honest works in the genre. The book is almost a foot square, a handsome photographic retrospective of his work with about 320 illustrations, a fourth of them in color.

Christo wraps and packs. He has wrapped bicycles, women, trees, office furniture, a medieval tower in Spoleto, storefronts and art museums. His wrapping is neither of the subtly appropriate kind found in Japanese folk art, such as the mat-wrapped trees in the gardens of the Kyoto palaces in winter, nor of the clamorous thermoplastic type that fills our supermarkets to form one medium for what is called pop art. Christo's packaging is always a handcraft performance: the half-enigmatic bulk is wrapped with cheap fabric or plastic film, then tied with cords and ropes as it seems anyone might do it. He exhibits and conceals at the same time, transforming texture and detail while suggesting the form within. There is a galvanizing strangeness in almost everything so treated. It is also plain that he has a sculptor's eye for curve and swell; not much is accidental.

Christo's packages are not merely symbolic. They need working fastenings, craftsmanship, scaffolds and finally, as their scale grows, engineering planning. His masterpiece (well illustrated in the last sixth of the book) was the wrapping of a mile-long stretch of rocky cliff along the beach of Botany Bay on the outskirts of Sydney, Australia. The fabric was a cheap off-white open-weave polypropylene, about a million square feet of it, fastened with 35 miles of polypropylene rope. A dozen professional rock-climbers worked on the cliff face, which was 80 feet high in places, unrolling the large fabric segments and stitching them together, and firing steel studs into the sandstone with Ramset guns. To these studs ropes were tied to hold the fabric firm. Students and artists, with some hired workmen, finished the wrapping in about three weeks—17,000 man-hours of labor.

The spectacle was fascinating. The old shapes of the rocky shore were now spatially filtered and entirely altered in color and surface. Even sea and sky appeared bluer; the world was somehow new and strange to the thousands of visitors who clambered around the magnificent stage that had arisen in Botany Bay. High winds claimed the fabric in six weeks; even its billowing and tattered dispersal to the open sea was a prolonged spectacle of form and motion. The finest record, indeed the only record, of the work is an excellent color motion-picture film and a sheaf of still photographs. The film closes, like the book but even more dramatically, with a long aerial shot of the wrapped cliff. The camera searches the shoreline as the plane slowly banks, looking for the work of Christo; finally you make it out. What was on the ground a remaking of nature on a gigantic scale, an entire wild landscape created by artifice, is from the air a small stripe, hard to see against an area of many square miles of beach and cliff. That single sharp evocation of scale is a convincing appraisal of the work.

It is easy to carp at such actions. Impermanent? Indeed they are, as is any movie location. Wasteful? Christo's wrapped cliff cost no more than any half-hour network television program, and a film record remains. Damaging to wildlife and to nature? The environment? Everything is gone now, except for a few studs that would be hard to find in the barren rock of the shore. The tidal flats, with their intricate ecosystem, were carefully avoided. The main result is that the residents of the city and many other people around the world know and feel something new about that shoreline.

In 1971 Christo tried one more grand

work. He proposed to stretch a steel cable across a valley in the Rockies and from that cable to hang a translucent divided curtain, parting the valley as a theatrical curtain parts the audience and the stage. As you walked or drove down the central highway you would pass through the curtain at its central parting and see the undivided view again. The valley Christo chose was near Rifle, Colo.; it called for a 1,400-foot cable between concrete anchors. The valley curtain, of an orange woven-synthetic fabric, would have weighed 25 tons; it would have been 365 feet high at the ends and would have hung 200 feet above the central road. Engineers were consulted on the structural design. In the event it was seen that the curtain fabric was not able to withstand the mountain winds, and the project was dropped halfway.

Christo will not delight everyone, but there is a case to be made for his having provided us with a new, playful, untrammeled view of technology, at once understanding and ironic, to help heal a world where the power to build seems so closely linked to cupidity, spoliation or terror.

INADVERTENT CLIMATE MODIFICATION. Report of the Study of Man's Impact on Climate (SMIC). The MIT Press ($12.50). Meeting as guests of the Swedish academies in a conference center near Stockholm during the summer of 1971, 30 atmospheric scientists from Paris and Ahmedabad, Tokyo, Leningrad and Boulder, and from nine other countries, worked to produce this prompt, terse and informed study in a single volume. The first quarter of the book is intended specifically to "provide an overview of the topic of inadvertent climate modification for the lay reader." The rest of the book, although still compact in form and sticking closely to the account of methods and results, implies a familiarity with quantitative argument in physics. The narrative is well documented with a very up-to-date set of references; it is an admirable source for study. The international list of authoritative participants and the consensus they reached leave the reader satisfied on all points save one, which is perhaps the most important: nearly all the recommendations are for more measurements and more theory! The candid and disarming refrain of the book is: "We don't know yet."

The background issues are nonetheless pretty clear. Our species lives in an unusual geologic period, cooler and more fluctuating than usual. Ever since slow drift brought the continents and the rotation axis of the earth to about their present position the middle latitudes have been cooling off. (The Tropics have stayed the same.) Ice ages mark the last million years or so; their cause is unclear. No plausible man-made changes can rival these natural ones; to estimate the results of what man does requires a general theory of how physical processes give rise to the present climate. Such a theory we do not have, although between complex computer models and improving data and experiment we are neither entirely ignorant nor without hope of powerful knowledge.

The most widespread impact of man on the climate seems to be much older than modern industry; it is the work of man the farmer. The dry forests of the semihumid Tropics have been gradually converted into savanna grassland across Africa and Brazil by the bush fires ignited in dry seasons, and well-developed forests in Europe, North America and Southwest Asia have been turned to arable land, more or less bare for a season. This has affected the heat and water losses over almost 20 percent of the earth's land area.

The second-largest effect is the increased evaporation that necessarily follows the irrigation of dry lands. On a local scale these effects on climate are already marked; it may be that the Rajputana Desert of India is man-made, and other regions might follow. Dust is natural as well as man-made; until now the natural sources of particulate matter, from volcanoes, winds and sea spray, seem to outweigh what we do. The carbon dioxide produced by man's fires is a clear example; it is not yet globally determining. The turbidity map of the U.S. shows that the maximum is still over the Great Smokies; it is a natural smog made by trees.

There are nonetheless triggers set and amplifiers waiting. The snow and ice surfaces are perhaps the most sensitive, particularly the arctic sea ice. An ice floe of the pack has an average lifetime of only a few years before it drifts toward the south to melt. The pack ice is unstable: once some ice melts the sea receives solar energy that the icy white surface would have reflected back into space and the water warms, so that more ice will melt. It seems possible that a large campaign to spread black dust on the ice from aircraft could succeed in removing most of the arctic ice at least temporarily by this positive feedback, with unknown consequences for the world's climate. (The removal of the floating ice would not of itself raise the sea level; only the melting of the high ice caps on Greenland or the Antarctic would do that. They too have a positive feedback mechanism: if the amount of ice decreases, so does the altitude of the ice cap, and the more ice there is at the warmer temperatures of lower altitudes, the faster the ice melts.

Hence the book makes a clear recommendation that "international agreement be sought to prevent large-scale (directly affecting over 1 million square kilometers) experiments in persistent or long-term climate modification." In other words, don't muck around with the bay! There are well-meant schemes of long standing to unfreeze the pack ice so as to improve—the promoters hope—the frozen lands around the Arctic. Experiment on a smaller scale seems indispensable.

Moscow sends out three times as much heat as it receives from the sun; New York radiates seven times as much. These are highly localized examples; man's total power is about the same as geothermal power, less than sunlight by a factor of 6,000. It is true that men build more furnaces each year. What else is the point of this forward-looking book? It seems timely enough; we have much to do to manage life prudently for 10^{10} humans on this planet. The ice pack is still, however, less vulnerable than the lives of the dispossessed and the weak.

THE JESUS BAG, by William H. Grier, M.D., and Price M. Cobbs, M.D. McGraw-Hill Book Company ($6.95). In 10 brief essays, strung with small case histories like beads of grace and of despair, two black psychiatrists from San Francisco diagnose and prescribe for the American society they view as Imperial Rome: "arrogant, vulnerable, and possibly on the verge of a mulish, obstinate death." Theirs is a productive anger; it organizes their energies but never leads them to self-deception, whether of history or of psychology. The pages glow with clear argument, phrased with deep concern and illustrated with palpable truths.

Their thesis can be summarized thus: Black people have suffered for three centuries in the hostile, murderous environment of white America. For them survival demanded inhibition of the normal reciprocal hostile responses. A protective structure of "commands of conscience and the strictures of religion"

arose, became overinhibitory and now curbs a range of "aggressive and competitive acts, most of which are healthy." That is *The Jesus Bag*. Yet now the moral riches of brotherhood latent in that experience, not simply historical Christianity but a creative graft onto the root, are available for a new American ethic: "the only way left for a nation so powerful it cannot be subjected to external control and so at war with itself that it cannot yield to intellect."

The Jesus Bag demands a full reading, but it may not be inappropriate to give three particulars drawn from many. The book's opening pages contain an explicit account of " 'the dozens,' " "a degrading, humiliating custom of mutual vilification practiced by black boys for as long as anyone can remember." These are essentially rhyming obscenities, ritualized, witty and combative, in which the speaker and most often the mother of the victim are described in some gross, outrageous sexual act. The authors see this trial as being socially adaptive, as a demand for a man to be stoical, as an insistence that although humiliation may be the texture of his life, his manhood can transcend even that.

Second, the authors celebrate the black therapist, a person who must be fully eclectic in his methods, an advocate and at the same time a healer judging his peers by the freedom they secure for their patients and not by their own militancy. Such a healer is "generally suspicious of contented men, but he is stirred to a therapeutic ardor by contented blacks, convinced as he is that a black man contented with his circumstances in America is very sick indeed." "How can one trust his mental life to one other than an honorable, courageous, moral man? Black patients are careful about the selection; they cannot afford mistakes. Whites might ask themselves if they, in all their affluence, can themselves afford a lesser man."

Third, there is a special bibliography, with some 600 citations from the world medical literature in English between the early 1950's and the middle 1960's, on the relation between hypertension and race. The list provides formidable support for the authors' remarks that the cause of the greater severity of high blood pressure among blacks is the active repression of hostile feelings, "being black, and perpetually angry, and unable to express it or do anything about it." Drs. Grier and Cobbs have done something weighty for us all; it is their readers who must act to end this American crime against Americans.

INDEX OF ADVERTISERS

JUNE 1972

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

MISSILE SUBMARINES AND NATIONAL SECURITY

POLARIS & POSEIDON, FBM FACTS. Strategic Systems Project Office. Navy Department, 1970.

SIPRI YEARBOOK OF WORLD ARMAMENTS AND DISARMAMENT 1969/70. Stockholm International Peace Research Institute. Humanities Press, 1971.

BEYOND SALT ONE. Herbert Scoville, Jr., in *Foreign Affairs,* Vol. 50, No. 3, pages 488–500; April, 1972.

REPORT ON ULMS. Members of Congress for Peace through Law. Washington, D.C., April, 1972.

TESTIMONY OF THE FEDERATION OF AMERICAN SCIENTISTS BEFORE DEFENSE SUBCOMMITTEE, APPROPRIATIONS COMMITTEE, U.S. House of Representatives on the FY 1973 Defense Program and Budget. Presented by Herbert Scoville. April, 1972.

MARKERS OF BIOLOGICAL INDIVIDUALITY

THE UNIQUENESS OF THE INDIVIDUAL. P. B. Medawar. Basic Books, Inc., 1957.

THE IMMUNOLOGY OF TRANSPLANTATION. P. B. Medawar in *The Harvey Lectures, Series LII (1956–1957).* Academic Press, 1958.

TRANSPLANTATION OF TISSUES AND CELLS. Edited by R. E. Billingham and Willys K. Silvers. The Wistar Institute Press, 1961.

TRANSPLANTATION ANTIGENS. Barry D. Kahan and Ralph A. Reisfeld in *Science,* Vol. 164, No. 3879, pages 514–521; May 2, 1969.

TRANSPLANTATION ANTIGENS. R. A. Reisfeld and B. D. Kahan in *Advances in Immunology,* Vol. 12, pages 117–200; 1970.

TRANSPLANTATION ANTIGENS: METHODS OF BIOLOGIC INDIVIDUALITY. Edited by Barry D. Kahan and Ralph A. Reisfeld. Academic Press, in press.

ORGANIC MATTER IN METEORITES

METEORITES. Brian Mason. John Wiley & Sons, Inc., 1962.

ORGANIC CONSTITUENTS OF METEORITES—A REVIEW. John M. Hayes in *Geochimica et Cosmochimica Acta,* Vol. 31, No. 9, pages 1395–1440; September, 1967.

CHEMICAL EVOLUTION. Melvin Calvin. Oxford University Press, 1969.

EVIDENCE FOR EXTRATERRESTRIAL AMINO-ACIDS AND HYDROCARBONS IN THE MURCHISON METEORITE. Keith Kvenvolden, James Lawless, Katherine Pering, Etta Peterson, Jose Flores, Cyril Ponnamperuma, I. R. Kaplan and Carleton Moore in *Nature,* Vol. 228, No. 5275, pages 923–926; December 5, 1970.

THE EVOLUTION OF REEFS

REVOLUTIONS IN THE HISTORY OF LIFE. Norman D. Newell in *Uniformity and Simplicity: A Symposium on the Principle of the Uniformity of Nature.* The Geological Society of America, Special Paper 89, edited by Claude C. Albritton, Jr.; 1967.

AN OUTLINE HISTORY OF TROPICAL ORGANIC REEFS. Norman D. Newell in *American Museum Novitates,* No. 2465; September 21, 1971.

REEF ORGANISMS THROUGH TIME. Proceedings of the North American Paleontological Convention, edited by Ellis Yochelson. Allen Press, Inc., 1971.

TEMPERATURE CONTROL IN FLYING MOTHS

TEMPERATURE REGULATION IN THE SPHINX MOTH, *MANDUCA SEXTA*: II, REGULATION OF HEAT LOSS BY CONTROL OF BLOOD CIRCULATION. Bernd Heinrich in *The Journal of Experimental Biology,* Vol. 54, No. 1, pages 153–166; February, 1971.

DEPENDENCE OF FLIGHT ON TEMPERATURE REGULATION IN THE MOTH, *MANDUCA SEXTA.* M. J. McCrea and J. E. Heath in *The Journal of Experimental Biology,* Vol. 54, No. 2, pages 415–435; April, 1971.

AN ANALYSIS OF PRE-FLIGHT WARM-UP IN THE SPHINX MOTH, *MANDUCA SEXTA.* Bernd Heinrich and George A. Bartholomew in *The Journal of Experimental Biology,* Vol. 55, No. 1, pages 223-239; August, 1971.

NONSTANDARD ANALYSIS

THE HISTORY OF THE CALCULUS AND ITS CONCEPTUAL DEVELOPMENT. Carl B. Boyer. Dover Publications, Inc., 1959.

NON-STANDARD ANALYSIS. Abraham Robinson. North-Holland Publishing Co., 1966.

NEW MODELS OF THE REAL-NUMBER LINE. Lynn Arthur Steen in *Scientific American,* Vol. 225, No. 2, pages 92–99; August, 1971.

INFINITESIMALS. A. H. Lightstone in *The American Mathematical Monthly,* Vol. 79, No. 3, pages 242–251; March, 1972.

CONTOUR AND CONTRAST

MACH BANDS: QUANTITATIVE STUDIES ON NEURAL NETWORKS IN THE RETINA. Floyd Ratliff. Holden-Day, Inc., 1965.

LINEAR SYSTEMS ANALYSIS OF THE *LIMULUS* RETINA. Frederick A. Dodge, Robert M. Shapley and Bruce W. Knight in *Behavioral Science,* Vol. 15, No. 1, pages 24–36; January, 1970.

CONTOUR AND CONTRAST. Floyd Ratliff in *Proceedings of the American Philosophical Society,* Vol. 115, No. 2, pages 150–163; April, 1971.

INHIBITORY INTERACTION IN THE RETINA OF *LIMULUS.* H. K. Hartline and F. Ratliff in *Handbook of Sensory Physiology: Vol. VII/Part IB.* Springer-Verlag, in print.

PSYCHOLOGICAL FACTORS IN STRESS AND DISEASE

SOMATIC EFFECTS OF PREDICTABLE AND UNPREDICTABLE SHOCK. Jay M. Weiss in *Psychosomatic Medicine,* Vol. 32, pages 397–408; 1970.

EXPERIMENTALLY INDUCED GASTRIC LESIONS: RESULTS AND IMPLICATIONS OF STUDIES IN ANIMALS. Robert Ader in *Advances in Psychosomatic Medicine,* Vol. 6, pages 1–39; 1971.

EFFECTS OF COPING BEHAVIOR IN DIFFERENT WARNING SIGNAL CONDITIONS ON STRESS PATHOLOGY IN RATS. Jay M. Weiss in *Journal of Comparative and Physiological Psychology,* Vol. 77, No. 1, pages 1–30; October, 1971.

MATHEMATICAL GAMES

PROBLEMS IN MODERN MATHEMATICS. Stanislaw M. Ulam. John Wiley & Sons, Inc., 1964.

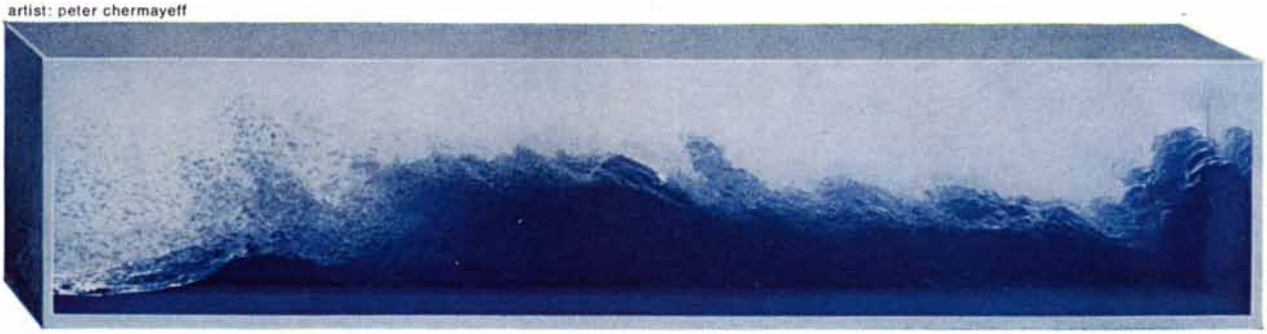
A SELECTION OF PROBLEMS IN THE THEORY OF NUMBERS. Wacław Sierpiński, translated by S. Sharma. The Macmillan Company, 1964.

TOMORROW'S MATH. C. Stanley Ogilvy. Oxford University Press, 1972.

THE AMATEUR SCIENTIST

TECHNIQUES OF OBSERVING THE WEATHER. B. C. Haynes. John Wiley & Sons, Inc., 1947.

artist: peter chermayeff



The ideal
Seas that are permanently protected from man's abuse.

The real
Thoughtlessly, man spews waste into the world's oceans. From the air, from the streams, from ships, all of it from ourselves.

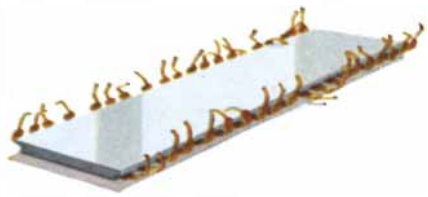
We must find new and better ways to guard our waters from ourselves. Our solutions must be swift. They must be creative and mature. For tomorrow the waters of the world will inherit what we do today.

Throughout the world, man must learn to function without fouling the oceans—and the air and earth that adjoin them. Until then, we cannot protect the environment in which life began—and on which our lives still depend.

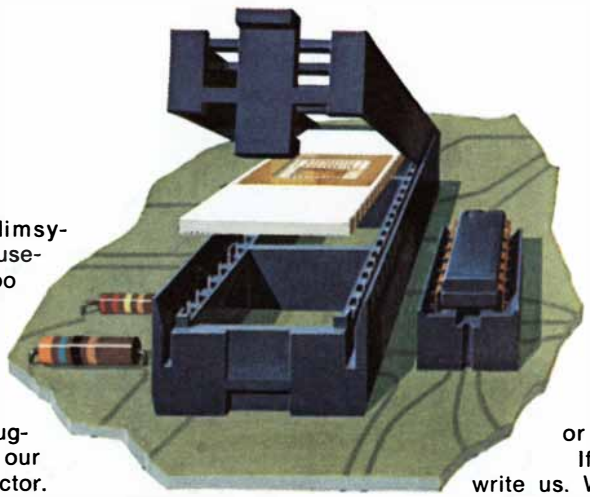
AtlanticRichfieldCompany

magnum: erich hartmann

The bug is dead.



Long live the Leadless Package Connector.



In MOS/LSI, the flimsy-limbed bug has outlived his usefullness. Forty legs are just too many to stick into a PC board, without bending or breaking. And one break costs you the whole package.

Now you can save that expense—and get package pluggability in the bargain—with our new, Leadless Package Connector. It's got legs, too. But they're on the connector, not the package.

This, plus zero insertion force, means you can plug and unplug a leadless package as often as you like, without wear or damage. And save even more because leadless packages cost less to make than their leggy counterparts. Since they don't need lead frames, brazing or plating. Or handling during bonding, or testing or packaging.

If you're bugged by broken legs, write us. We'll send you some fact-filled information bulletins.

AMP Incorporated, Industrial Division, Harrisburg, Pa. 17105.

AMP
INCORPORATED

Manufacturing and direct sales facilities worldwide: Barcelona, Brussels, Buenos Aires, Frankfurt, London, Mexico City, Paris, Puerto Rico, Sao Paulo, s'Hertogenbosch (Holland), Sydney, Stockholm, Tokyo, Toronto, Turin, Vienna.